MAKROKOGNITIVE MODELLIERUNG MENTALER BEANSPRUCHUNG VON TOWERLOTSEN MIT FARBIGEN PETRINETZEN

vorgelegt von

Master of Science
Hardy Smieszek
aus Berlin

von der Fakultät V
Verkehrs- und Maschinensysteme
der Technischen Universität Berlin
im DFG-Graduiertenkolleg
Prospektive Gestaltung von Mensch-Technik-Interaktion
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
— Dr.-Ing. —

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. phil. Manfred Thüring
Gutachter: Prof. Dr.-Ing. Thomas Jürgensohn
Gutachter: Prof. Dr.-Ing. Nele Rußwinkel

Tag der wissenschaftlichen Aussprache: 24. Oktober 2014

Berlin, 2014

D 83

DANKSAGUNG

An erster Stelle möchte ich der Deutschen Forschungsgemeinschaft (DFG) für die Förderung dieser Arbeit im Rahmen des Graduiertenkollegs prometei an der Technischen Universität Berlin danken. Ferner geht der Dank an die TU Berlin, die das Abschließen dieser Dissertation durch die Förderung mit einem Durchstarterstipendium ermöglicht hat.

Besonderer Dank gilt vor allem meinen beiden Betreuern, Prof. Dr.-Ing. Thomas Jürgensohn und Prof. Dr.-Ing. Nele Rußwinkel von der Technischen Universität Berlin, die mir stets mit Ihrem Rat, regen Diskussionen und kritischen Nachfragen zur Seite standen und die Arbeit damit inhaltlich und methodisch bereichert und nicht zuletzt in ihrer Struktur vorangebracht haben.

Ferner bedanke ich mich bei Prof. Dr. Manzey für die Übernahme der Betreuung in den ersten zwei Jahren meiner Dissertation, sowie die weiteren methodischen Hinweise hinsichtlich der Erhebung physiologischer Daten.

Ein besonders herzlicher Dank geht an Christoph Möhlenbrink, Peer Manske und Andreas Hasselberg vom Deutschen Zentrum für Luft- und Raumfahrt (DLR) in Braunschweig für die fruchtbare Zusammenarbeit während meines Forschungsaufenthaltes in Braunschweig, aus der auch mehrere gemeinsame Veröffentlichungen entstanden, und denen ich darüber hinaus einen enormen Wissenszuwachs in Sachen Petrinetz-Modellierung zu verdanken habe. Speziell danke ich Christoph Möhlenbrink für seine konstruktiven Hinweise im Hinblick auf die thematische Schwerpunktsetzung und methodische Ausrichtung der Arbeit. Ohne ihn wäre die Arbeit nicht da, wo sie heut ist.

Weiterer Dank geht an Henk Blom vom National Aerospace Laboratory (NLR) in Amsterdam für seine aufopferungsvolle Suche nach der verschollen geglaubten Aufgabenanalyse von Cox (1994), sowie Erik Hollnagel, für seine inspirierenden Veröffentlichungen, sowie seine kritischen Hinweise zur Modellierung in den Anfängen dieser Arbeit.

Ferner wäre die Arbeit ohne die Mitarbeit meiner wissenschaftlichen Hilfskräfte und Studenten, die mich aufopfernd bei meinen Recherchen und Studien unterstützten und Arbeiten im Themenfeld anfertigten, nicht möglich gewesen. Großer Dank geht in chronologischer Reihenfolge an Juliane Bock, Toralf Stubbe, Christian Karl, Dominik Sittel und Fabian Joeres.

Darüber hinaus möchte ich den zahlreichen Fachleuten meinen Dank aussprechen, ohne deren Expertise, Hilfe und Einschätzung ich sicherlich an der einen oder anderen Stelle etwas verloren gewesen wäre. Dazu zählen vor allem die Belegschaft des Towers in Schönefeld und deren Supervisor Hendrik Finger, sowie die Belegschaft des Towers Braunschweig-Wolfsburg, die mir die Möglichkeit gaben, die Arbeitsumgebung und Arbeitsaufgaben von Towerlotsen live

mitzuerleben und Fragen zu stellen. Darüber hinaus danke ich Herrn Ferdinand Behrendt vom Institut für Luft- und Raumfahrt der Technischen Universität Berlin für die kompetente Beantwortung unserer Fragen und die Herstellung des Kontaktes zum Tower in Tegel.

Gespräche in den Arbeitsgemeinschaften und Ringvorlesungen des Graduiertenkollegs, mit den Kollegen des Graduiertenkollegs, sowie des Fachgebietes „Kognitive Modellierung in dynamischen Mensch-Maschine-Systemen“ haben die Arbeit sehr vorangebracht. So danke ich nicht nur für den inhaltlichen und methodischen Rat stellvertretend für das gesamte Kolleg und das Fachgebiet Fabian Joeres, Maria Luz, Michael Minge, Janna Protzak, Stefan Schaffer, Enrico Maier, Sebastian Werk, Stefan Lindner, Christian Stößel, Elisabeth Brandenburg, Matthias Zickerow und Sabine Prezenski.

Besonders hervorheben möchte ich meine Freundin Stefanie Huber, die nur zu gut die Hochs und Tiefs, die Leidenswege und Euphorie Phasen einer Dissertation kennt und daher auch immer erste Anlaufstelle für seelische Beistandssuche meinerseits war, wenn das tägliche auf- und ab wieder einen Tiefststand erreicht hatte. Doch nicht nur durch ihren privaten Beistand und ihren Mut-Zuspruch, sondern auch fachlich hat sie diese Arbeit bereichert und maßgeblich nach vorn gebracht. Ohne sie hätte ich die Arbeit wohl nie, oder zumindest nicht so schnell und motiviert zu Ende gebracht.

Dank gilt auch dem Administratoren-Team des GRK prometei, Steffen Hartwig, Nicolaus Rötting, sowie Sebastian Lohmeier, die bei Problemen immer zur Stelle waren und den Rechner wieder zum Laufen brachten.

Ich danke auch meiner guten Freundin Ulrike Schmuntzsch, die zur gleichen Zeit an ihrer Dissertation arbeitete wie ich, für die aufschlussreichen Diskussionen beim Mittagessen und die Möglichkeit, sich gemeinsam über die Unzufriedenheit mit der Gesamtsituation Luft verschaffen zu können. Meinen Freunden Alexander Schulze, Susann Sträßer, Katja Fischer, Manuel Haupt, Denise Wocko und Benjamin Kania danke ich für die erheiternden Stunden, ihren Beistand und Zuspruch.

Nicht zuletzt danke ich meiner Familie, sowohl den Smieszeks als auch den Meins, für ihren unermüdlichen Glauben und die oft gesagten Worte „du schaffst das!“

KURZFASSUNG

In den Vergangenen Jahren hat der Flugverkehr weltweit stark zugenommen und wird auch in den kommenden Jahren weiter zunehmen. Hauptaufgabe der Lotsen am Flughafen ist die sichere und flüssige Abwicklung des Boden- und Luftverkehrs innerhalb ihrer Kontrollzone. Dabei sind die Lotsen einem ständigen trade-off zwischen Effizienz und Gründlichkeit ausgesetzt, welcher bei steigendem Flugverkehrsaufkommen Richtung Effizienz verschoben werden muss, wodurch das Risiko erhöht wird, dass vom Lotsen Fehler gemacht werden und es damit zu kritischen Situationen auf dem Rollfeld kommt. Ist bekannt, zu welchen Zeitpunkten ein Lotse Gefahr läuft überlastet zu werden, kann gegengesteuert werden, um beispielsweise kognitive Kapazitäten freizugeben und das Risiko für Fehler zu senken. Es ist daher von größter Wichtigkeit, den Verlauf der mentalen Beanspruchung des Lotsen über die Zeit zu untersuchen, zu beobachten und im Idealfall sogar vorhersagen zu können.

Die zur Aufnahme mentaler Beanspruchung vorhandenen Methoden sind allerdings für deren Erfassung und Vorhersage nicht vollständig geeignet. Auch der Versuch mentale Beanspruchung von Fluglotsen mittels mathematischer Indizes vorherzusagen ist unzureichend, da lediglich objektive Belastungsfaktoren, wie das Verkehrsaufkommen, nicht aber subjektive Faktoren der Person, wie bspw. Erfahrung und Training, berücksichtigt werden. Mit Hilfe der kognitiven Modellierung soll diese Lücke geschlossen werden, wobei zur Lösung dieses Problems ein innovativer Modellierungsansatz erforderlich ist, welcher nicht nur die Umgebungsfaktoren in Betracht zieht, sondern auch das menschliche Handeln und die kognitiven Prozesse betrachtet. Durch die Simulation von Modellen ist dann eine Antizipation zukünftiger Situationen möglich.

Anhand der Richtlinien für die Entwicklung kognitiver Simulationen nach Cacciabue wurden einschlägige Analysen des Flugverkehrskontrollsystems sowie kognitionspsychologischer Theorien der Informationsverarbeitung und mentaler Beanspruchung durchgeführt, die als Grundlage für die Implementierung eines makrokognitiven Modells der Flughafen-verkehrskontrolle (MATriCS; Model of Airport Traffic Control System) auf der Basis farbiger Petrinetze dienen, welches zur Vorhersage der mentalen Beanspruchung von Towerlotsen verwendet wird. Die Validierung des Modells erfolgte in einer realitätsnahen Computersimulation der Towerlotsentätigkeit, wobei verschiedene Beanspruchungsmaße erhoben wurden (subjektive, physiologische, Leistungsmaße). Die Ergebnisse konnten zeigen, dass das Modell mentale Beanspruchung hinreichend diagnostisch abbildet und in der Lage ist sowohl den Trend der individuellen Beanspruchungsverläufe mit mittlerer Güte als auch die mittlere Beanspruchungshöhe zu replizieren.

ABSTRACT

In recent years air traffic has increased rapidly and will continue to increase in the next years. The main task of air traffic controllers at airports is the safe and efficient handling of aircraft of air and ground traffic within their designated control zone. In doing so they have to reach a tradeoff between efficiency and thoroughness which under the condition of growing air traffic has to be shifted towards efficiency whereby the risk of controller errors and critical situations on the airfield raises. It is therefore necessary to know at which point in time the controller is likely to get overloaded to be able to efficiently counteract in order to release cognitive capacity and therefore to reduce the risk of error. It is of great importance to investigate, observe and ideally to predict the distribution of mental workload of the controller over time.

The available methods for capturing mental workload are, however, not fully appropriate for its detection and prediction. The attempt to predict mental workload of air traffic controllers by using mathematical indices is insufficient as well, since solely objective factors of load like traffic amount but not subjective factors like training or experience are taken into account. With the help of cognitive modelling this gap shall be closed. In order to solve this problem an innovative approach to modelling is needed which not only takes environmental factors but also human behaviour and cognitive processes into account. Due to simulation of models an anticipation of future situations is possible.

Following basic guidelines for the development of cognitive simulation according to Cacciabue relevant analyses of the air traffic control system as well as theories of information processing and mental workload from cognitive psychology were conducted which serve as a foundation for implementing of a macro-cognitive model of airport traffic control system (MATriCS; Model of Airport Traffic Control System) based on coloured petri nets which is used for prediction of mental workload of tower controllers. The validation of the model took place within a realistic computer simulation of the tasks of tower controllers while different measures of mental workload were gathered (subjective, physiological, performance measures). Results show that the model illustrates mental workload with sufficient diagnosticity and is able to replicate both the trend of the individual distribution of mental workload over time with average goodness as well as the mean workload level.

INHALTSVERZEICHNIS

1 Einleitung 2

1.1 Problemstellung 2

1.2 Reichweite der Arbeit 4

1.3 Forschungsziele 4

1.4 Aufbau der Arbeit 6

2 Mentale Beanspruchung, Kognitive Kapazität und Arbeitsgedächtnis 8

2.1 Mentale Beanspruchung – Theorien und Modelle 8

2.1.1 Aktivierungstheoretische Modelle 9

2.1.2 Aufmerksamkeitstheoretische Modelle 10

2.2 Arbeitsgedächtnismodelle, Kapazität und Chunks 13

2.3 Mentale Beanspruchung in der Flugverkehrskontrolle 17

2.4 Erfassung Mentaler Beanspruchung und Operateurszustände 22

2.4.1 Physiologische Maße 23

2.4.2 Subjektive Maße 25

2.4.3 Leistungsbasierte Maße 27

2.4.4 Zusammenfassende Bewertung der Erfassungsmethoden 28

2.5 Kapitelzusammenfassung 30

3 Stand der Forschung 31

3.1 Modellierung und Simulation von Mensch-Maschine-Systemen 31

3.2 Mikro- und Makrokognitive Forschungs- und Modellierungsansätze 34

3.2.1 Mikrokognitive Ansätze und kognitive Architekturen 34

3.2.2 Makrokognitive Ansätze 36

3.2.3 Die Verbindung Mikro- und Makrokognitiver Ansätze 38

3.3 Ausgewählte Modelle und Simulationen von Problemstellungen im Mensch-Maschine-System Luftfahrt und Flugverkehrskontrolle 41

3.4 Kognitive Modellierung Mentaler Beanspruchung 45

3.5 Kapitelzusammenfassung 47
4 Entwicklung eines makrokognitiven Modells der Flughafen-verkehrskontrolle (MATriCS) 49
4.1 Vorgehensweise zur Entwicklung Kognitiver Simulationen 49
4.2 Definition der Problemgrenzen 52
4.3 Aufgabenanalyse 56
4.3.1 Aufgaben, Verantwortlichkeiten, Verpflichtungen 58
4.3.2 Beschreibung der Aufgabenanalysen 59
4.3.3 Analyse aus dem EUROCONTROL-Projekt MANTEA 59
4.3.4 Analyse aus dem EUROCONTROL-Projekt ATHOS 60
4.3.5 Analyse der FAA (FAA Air Traffic Control Operations Concepts) 60
4.3.6 Analyse aus dem DFS-Projekt VICTOR 61
4.3.7 Analyse des Royal Air Force Institute of Aviation Medicine 62
4.4 Feldstudie der Arbeitsumgebung 64
4.4.1 Allgemeine Abläufe und Zuständigkeiten 64
4.4.2 Hilfsmittel zur Informationsaufnahme und Aufgabenbearbeitung 66
4.4.3 Position Towerlotse (TWR) 68
4.4.4 Abwicklungsprioritäten und Staffelungsminima des TWR 69
4.4.5 Position Rollverkehrslotse (Ground; GND) 70
4.4.6 Prioritäten des GND 70
4.5 Einordnung in und Erstellung des theoretischen Modells 71
4.6 Auswahl des Beschreibungsmittels 76
4.7 Kapitelzusammenfassung 82
5 Beschreibung und Implementierung des entwickelten Modells 84
5.1 Allgemeine Anmerkungen zum Modell 84
5.2 Das Flughafenmodell 87
5.2.1 Theoretische Erläuterungen 87
5.2.2 Das Petrinetz-Modell im Detail 88
5.2.3 Colorsets und einlesen der Initialen Markierung 93

5.3 Das Interaktionsmodell ... 95
5.3.1 Theoretische Erläuterung ... 95
5.3.2 Das Petrinetz-Modell im Detail ... 96
5.4 Das Lotsenmodell ... 99
5.4.1 Theoretische Erläuterungen ... 99
5.4.2 Phase der Informationsaufnahme ... 101
5.4.3 Das perceive-Modell im Petrinetz ... 102
5.4.4 Phase der Planung ... 108
5.4.5 Das plan-Modell im Petrinetz ... 109
5.4.6 Phase der Entscheidungsfindung ... 110
5.4.7 Das Entscheidungsmodell im Petrinetz ... 114
5.4.8 Phase der Handlungsausführung ... 118
5.4.9 Das Handlungs-Modell im Petrinetz ... 121
5.4.10 Das Arbeitsgedächtnismodell ... 125
5.4.11 Arbeitsgedächtnis im Petrinetzmodell ... 127
5.5 Simulation und Analyse des Petrinetzmodells ... 131
5.6 Kapitelzusammenfassung ... 132
6 Validierung ... 134
6.1 Methodik ... 134
6.1.1 Teilnehmer/Stichprobe ... 135
6.1.2 Simulator und Material ... 136
6.1.3 Versuchsdesign ... 138
6.1.4 Hypothesen ... 143
6.1.5 Versuchsablauf ... 156
6.2 Ergebnisse ... 159
6.2.1 Zusammenhang d2-R Ergebnis und Fehlerzahl bzw. Höhe der Beanspruchungsratings (Hypothese 1) ... 159
6.2.2 Bewertung der Sensitivität der Herzrate (Hypothese 2) ... 162

6.2.3 Bewertung der mittleren mentalen Beanspruchung mittels Herzrate (Hypothesen 3a-f) ... 163
6.2.4 Bewertung der mittleren mentalen Beanspruchung mittels RSME (Hypothesen 4a – f) ... 164
6.2.5 Bewertung der mittleren mentalen Beanspruchung mittels NASA-TLX Dimension geistige Anstrengung (Hypothesen 5a – f) ... 166
6.2.6 Fehlerzahl (Hypothesen 6a – f) ... 168
6.2.7 Mittlere Beanspruchungshöhe des Modells (Hypothese 7) ... 169
6.2.8 Trendvergleich der mittleren Beanspruchungshöhe (Hypothese 8) ... 171
6.2.9 Überprüfung der Modellgüte mittels Einzelfallanalysen (hypothese 9) ... 173
6.2.10 Explorative Analyse: Absolute Abweichung der Modelldaten ... 177
6.2.11 Explorative Analyse: Fehler und Beanspruchungspeaks ... 178
6.3 Diskussion der Ergebnisse ... 179
6.3.1 Zusammenhang d2-R Ergebnis und Fehlerzahl bzw. Höhe der Beanspruchungsratings (Hypothese 1) ... 179
6.3.2 Bewertung der Sensitivität der Herzrate (Hypothese 2) ... 180
6.3.3 Bewertung der mittleren mentalen Beanspruchung erfasst durch die RSME-SKALA (Hypothese 4) ... 182
6.3.4 Bewertung der Mittleren mentalen Beanspruchung erfasst durch den NASA-TLX (Hypothese 5) ... 182
6.3.5 Bewertung der mittleren Fehlerzahl (Hypothese 6) ... 184
6.3.6 Mittlere Beanspruchungshöhe der Modelldaten (Hypothese 7) ... 185
6.3.7 Trendvergleich der durch subjektive und leistungsbasierte Methoden erfassten Beanspruchungshöhe mit dem Modell (Hypothese 8) ... 186
6.3.8 Überprüfung der Modellgüte mittels Einzelfallanalysen (Hypothese 9) ... 186
6.3.9 Explorative Analyse: Absolute Abweichung der Modelldaten ... 189
6.3.10 Explorative Analyse: Fehler und Beanspruchungspeaks ... 189
6.4 Kapitelzusammenfassung ... 190
7 Diskussion der Arbeit ... 193

7.1 Beitrag zur Ableitung einer geeigneten Modellierungsform des Konstruktes mentaler Beanspruchung (Forschungsziel 1) ... 193

7.2 Beitrag zur Verfolgung eines makrokognitiven Modellierungsansatzes und Verbindung mit mikrokognitiven Ansätzen (Forschungsziel 2) ... 195

7.3 Beitrag hinsichtlich der Erstellung eines Modells zur quantitativen Beschreibung mentaler Beanspruchung von Towerlotsen in Abhängigkeit verschiedener Einflussfaktoren (Forschungsziel 3) ... 197

7.4 Beitrag hinsichtlich der Validierung des erstellten Modellansatzes mittels statistischer Methoden (Forschungsziel 4) ... 200

8 Ausblick ... 204

Literaturverzeichnis ... 207

ABBILDUNGSVERZEICHNIS

Abbildung 1: Kognitives Model der Fluglotsenaufgabe (eigene Darstellung nach Koros el al. 2006; Mogford et al. 1995) 19

Abbildung 2: Zusammenhang zwischen Eingangsfaktoren, mentaler Beanspruchung und direkt messbaren Variablen (eigene Darstellung nach Hollnagel & Woods, 2005) 21

Abbildung 3: Übersicht der Ansätze und zugehöriger Werkzeuge zur Modellierung von Mensch-Maschine-Systemen (eigene Darstellung) 38

Abbildung 4: Verbindung zwischen makro-, mikrokognitiver und neuronaler Ebene durch Up- und Downscaling (eigene Darstellung nach West et al. 2013). 39

Abbildung 5: Richtlinien für die Entwicklung kognitiver Simulationen modifiziert nach Cacciabue (1998b) (eigene Darstellung) 50

Abbildung 6: Dreiteilung von Mensch-Maschine-Modellen (eigene Darstellung nach Cacciabue, 1998a) 53

Abbildung 7: Phasen des Finalen Anfluges und die dazugehörigen Aufgaben des Towerlotsen (eigene Darstellung nach Tavanti, 2006b) 65

Abbildung 8: Der Kontrolltower Berlin Tegel von innen (Bildquelle: DFS; http://www.dfs.de/dfs_homepage/mediathek/de/) 66

Abbildung 9: Beispiel für einen Flugstreifen (eigene Darstellung) 67

Abbildung 10: Aufgaben von Towerlotsen eingebettet in die Phasen der Handlungsregulation nach Hacker (1986) (nach Smieszek, Huber & Jürgensohn, 2011) 72

Abbildung 11: Zuordnung der acht Hauptaufgaben des Rollverkehrslotsen (TWR) nach Cox (1994b) zu den Phasen der Handlungsregulation nach Hacker (1986) (nach Smieszek, 2011) ... 73

Abbildung 12: Unteraufgaben zur Ausführung der Hauptaufgabe 3 des Rollverkehrslotsen "Führen der Ankünfte zum Stand" (vgl. Smieszek, 2011) 75

Abbildung 13: Beispiel für ein einfaches Petrinetz 78

Abbildung 14: Beispiel für ein einfaches farbiges Petrinetz für einen Landeprozess 79

Abbildung 15: Das Petrinetz aus Abbildung 14 nach einmaligem feuern der Transition "Landing" 80

Abbildung 16: Beispiel für eine Substitutions-Transition in einem hierarchischen Petrinetz 80

Abbildung 17: Darstellung der drei Teile des Petri-Netz-Modells (Flughafen-, Interaktions- und Lotsenmodell) und deren Verbindungen (nach Smieszek et al., 2013) 85

Abbildung 18: Schematische Darstellung der hierarchischen Stuktur des MATriCS-Modells..... 86

Abbildung 19: Elemente der invariablen Prozesslogik (eigene Darstellung nach Manske, Smieszek, Hasselberg & Möhlenbrink, 2013)..... 87

Abbildung 20: Layout eines exemplarischen Flughafens (eigene Darstellung)..... 88

Abbildung 21: Oberste Ebene des Flughafenmodells 89

Abbildung 22: Einzelne Teile eines exemplarischen Flughafens (eigene Darstellung nach Manske, 2013)..... 90

Abbildung 23: Umsetzung der Substitutionstransition Taxiing (Rollen) im Petrinetz..... 91

Abbildung 24: Verbindungen zwischen den einzelnen Teilen eines exemplarischen Flugplatzes (eigene Darstellung nach Manske, 2013)..... 92

Abbildung 25: Das Interaction-Modell im Petrinetz..... 97

Abbildung 26: Das Lotsenmodell im Petrinetz 100

Abbildung 27: Das Devided-Sub-Modul im Petrinetz 103

Abbildung 28: Das Perceive-Sub-Modul im Petrinetz..... 104

Abbildung 29: Das focused-Sub-Modul im Petrinetz 106

Abbildung 30: Das Planungsmodul im Petrinetz 109

Abbildung 31: Das Entscheidungs-Modul im Petrinetz 113

Abbildung 32: Entscheidungs-Sub-Modul für Starts- und Landungen im Petrinetz..... 116

Abbildung 33: Relative Zielbreite (eigene Darstellung nach Preim und Dachselt, 2010) 119

Abbildung 34: Koordinatensystem am Lotsenarbeitsplatz..... 120

Abbildung 35: Das Act-Modul im Petrinetz..... 122

Abbildung 36: Das Manual-Modul im Petrinetz 124

Abbildung 37: Das Working-Memory-Modul (WM) im Petrinetz..... 126

Abbildung 38: Das WM_Storage-Modul im Petrinetz..... 128

Abbildung 39: Der WM_retrieval-Prozess im Petrinetz..... 131

Abbildung 40: Ausgabe des WM_ 131

Abbildung 41: Ausgabe des Requesting_TWR-Monitors in einer Textdatei..... 132

Abbildung 42: Arbeitsplatz der Versuchspersonen 137

Abbildung 43: Arbeitsplatz der Versuchspersonen mit den zur Verfügung stehenden Radarbildschirmen (links Luftradar, rechts Bodenradar) 157

Abbildung 44: Die den Versuchspersonen zur Verfügung stehende Stripbay mit einigen Flugstreifen 158

Abbildung 45: Mittlere Fehlerzahl der drei Gruppen nach ihrem Abschneiden im d2-R-Test ... 160

Abbildung 46: Mittlere Beanspruchungshöhe (laut RSME) der drei Gruppen nach d2-R-Ergebniss 160

Abbildung 47: Mittlere Beanspruchungshöhe (laut NASA-TLX g. A.) der drei Gruppen nach d2-R-Ergebniss 161

Abbildung 48: Mittlere IBI (in ms) der beiden Experimentalbedingungen (hohe und geringe Belastung) und der Baselinebedingung 162

Abbildung 49: Mittlere Beanspruchungshöhe jedes Szenarios (in %) erfasst mit dem RSME 164

Abbildung 50: Mittleren Beanspruchungshöhe jedes Szenarios (in %) erfasst mit dem NASA-TLX (Dimension geistige Anstrengung) 166

Abbildung 51: Mittlere Gesamtfehlerzahl pro Szenario 168

Abbildung 52: Mittleren Beanspruchungshöhe jedes Szenarios (in %) des Modells 170

Abbildung 53: Mittlere Beanspruchungshöhe des Modells (Anzahl Chunks) verglichen mit der mittleren Beanspruchungshöhe des RSME, des NASA-TLX (g.A.), sowie des NASA-RTLX (jeweils in %) 172

Abbildung 54: Streudiagramm Modell-RSME der Versuchsperson 24 im Szenario 1 174

Abbildung 55: Zeitliche Verläufe der subjektiven Beanspruchungswerte (blau) sowie der Modellwerte (rot) der Versuchsperson 24 im Szenario 1 174

TABELLENVERZEICHNIS

Tabelle 1: Vergleich der vorgestellten Methoden zur Erfassung mentaler Beanspruchung und Bewertung der Gütekriterien (++ sehr gut; -- sehr schlecht) 29

Tabelle 2: Bewegungstypen, die damit verbundene Anzahl der Features und die sich daraus ergebende Vorbereitungszeit für die definierten Ziel-Locations 121

Tabelle 3: Übersicht über die Szenarienbezeichnung und der mit ihnen einhergehenden Belastung 140

Tabelle 4: Permutation der Szenarien für die 24 Versuchspersonen 140

Tabelle 5: Übersicht der Hypothesen zur Überprüfung der experimentellen 148

Tabelle 6: Übersicht der Hypothesen zur Überprüfung der Modellgüte 155

Tabelle 7: Ergebnisse der Testung der Hypothese 1 161

Tabelle 8: Ergebnisse der Testung der Hypothese 2 163

Tabelle 9: Ergebnisse der Testung der Alternativhypothesen von Hypothese 4 165

Tabelle 10: Ergebnisse der Testung der Nullhypothesen von Hypothese 4 165

Tabelle 11: Ergebnisse der Testung der Alternativhypothesen von Hypothese 5 167

Tabelle 12: Ergebnisse der Testung der Nullhypothesen von Hypothese 5 167

Tabelle 13: Ergebnisse der Testung der Alternativhypothesen von Hypothese 6 169

Tabelle 14: Ergebnisse der Testung der Nullhypothesen von Hypothese 6 169

Tabelle 15: Ergebnisse der Testung der Alternativhypothesen von Hypothese 7 170

Tabelle 16: Ergebnisse der Testung der Nullhypothesen von Hypothese 7 171

Tabelle 17: Übersicht der Korrelationskoeffizienten der einzelnen Messskalenvergleiche 173

Tabelle 18: Zusammenfassung aller möglichen Korrelationen zwischen Modell und RSME pro Zeitreihe des Szenarios 1 175

Tabelle 19: Zusammenfassung der möglichen und positiven Korrelationen pro Szenario 176

Tabelle 20: Mittlere Korrelationen aller Zeitreihen pro Szenario, sowie über alle Szenarien 176

Tabelle 21: Absolute und mittlere Absolute Abweichung der Modelldaten von den subjektiven Daten (RSME und NASA-TLX) 177

Tabelle 22: Vierfeldertafel der Signalentdeckungstheorie bezogen auf die Fehlerzahl in Versuch und Modell 178

Tabelle 23: Zusammenfassung der Ergebnisse: Fehler innerhalb von Beanspruchungsgipfeln 179

„We often fail to realize how little we know about a thing until we attempt to simulate it on a computer."

Donald E. Knuth (1997, S. 298)

1 EINLEITUNG

Fluglotsen im Flughafentower haben die Aufgabe die Flugbewegungen am und um den Flughafen sicher und flüssig abzuwickeln und Zusammenstöße am Boden und in der Luft zu vermeiden. Ohne die Erlaubnis eines Lotsen (sogenannte Freigaben), darf kein Flugzeug oder Fahrzeug eine Aktion auf dem Flughafen ausführen. Dabei besteht die wichtige Aufgabe des Lotsen darin, zu gewährleisten, dass alle Bewegungen am Flughafen möglichst schnell und unmittelbar (mit dem Ziel der flüssigen und effizienten Abwicklung) von A nach B gelangen, ohne dass andere Verkehrsteilnehmer behindert werden oder es gar zu gefährlichen Annäherungen oder Zusammenstößen kommt (mit dem Ziel der sicheren und gründlichen Abwicklung). Sie sind dabei einem ständigen Trade-off zwischen Effizienz und Gründlichkeit ausgesetzt (ETTO-Principle; Efficiency-Thorougness-Trade-off; Hollnagel, 2009). Dies stellt eine hochkomplexe und dynamische Aufgabe dar, die ein hohes Maß an Wissen und Konzentration des Lotsen erfordert. Da die menschliche Informationsverarbeitungskapazität begrenzt ist, gilt es, diese Grenzen möglichst genau zu kennen um möglichst gut mit ihnen umgehen zu können. Sowohl eine Überlastung der Kapazität, als auch eine langfristige Unterforderung kann zu Fehlern führen. Hierbei stellt die steigende Zahl an Flügen allgemein und damit auch die steigende Zahl an Starts und Landungen, sowie Rollvorgängen (DFS, 2013), die die Towerlotsen koordinieren müssen, ein großes Problem dar, da die Lotsen ihr Handeln in Richtung des Effizienzzieles verschieben müssen und hierdurch Überlastungssituationen und damit Fehler wahrscheinlicher werden (Hollnagel, 2009). Zusätzlich führen Bestrebungen die Kontrolle über mehrere kleinere Flughäfen auf einen Lotsen zu übertragen (sogenannte Remote Control) zu einer indirekten Erhöhung der Flugzahlen (z.B. Fürstenau et al., 2009; Oehme & Schulz-Rückert, 2010).

Es ist daher von großer Bedeutung, Zeitpunkte identifizieren zu können, zu denen eine Überlastung des Lotsen wahrscheinlich ist, um effektiv gegensteuern zu können, was beispielsweise durch den Einsatz adaptiver Automations- und Unterstützungssysteme gewährleistet werden könnte, wie es vor allem in der Fahrzeug- und Flugzeugführung teilweise schon getan wird (vgl. bspw. Flemisch et al., 2012). Hierdurch können kognitive Kapazitäten des Lotsen mit dem Ziel freigegeben werden, dass dieser die Abwicklung der Flugzeuge mit höherer Effizienz und gleichbleibender Gründlichkeit nachkommen kann.

1.1 PROBLEMSTELLUNG

In der Flughafenverkehrskontrolle ist es dabei vor dem Hintergrund steigender Flugzahlen besonders sinnvoll, die Adaption einer Automation in Abhängigkeit der mentalen Beanspruchung des Lotsen vorzunehmen. So können unter Bedingungen hoher Belastung und

Beanspruchung des Lotsen, Funktionen an die Automation abgegeben werden, welche unter geringerer Belastung und Beanspruchung wieder dem Lotsen zufallen. Hierbei besteht allerdings das Problem, dass es sich bei mentaler Beanspruchung um eine „innere“ Variable des Menschen handelt und diese nicht direkt beobachtbar oder gar messbar ist. In der Mensch-Maschine-Systemtechnik und der Human Factors-Forschung besteht eine lange Tradition zur Erfassung mentaler Beanspruchung: Es existieren zahlreiche Methoden und Theorien darüber, wie durch die Erfassung direkt messbarer Parameter ein Rückschluss auf die mentale Beanspruchung von Operateuren ermöglicht wird. Dennoch können auch diese Erfassungsmethoden bis heute nicht als ausreichend angesehen werden, um beispielsweise im Hinblick auf die Gestaltung adaptiver Automationen die nötige Sensitivität und Diagnostizität zur Erkennung mentaler Beanspruchung aufzubringen. Es ist daher notwendig, alternative Ansätze zu finden, um diese Nachteile überwinden zu können. Hierbei stellt die kognitive Modellierung und Simulation von Mensch-Maschine-Systemen einen vielversprechenden Ansatz dar. Die Gestaltung adaptiver Automation erfolgte lange Jahre auf Grundlage von Expertensystemen[1] (z.B. Ruckdeschel, 1997), bedient sich allerdings mehr und mehr der kognitiven Modellierung (z.B. Svenmarck & Dekker, 2003), da hierbei nicht nur die bloßen Aufgabenabläufe, sondern auch die kognitiven Prozesse des Operateurs modelliert und simuliert werden können und dies zu einer höheren Korrespondenz zwischen Modell und Operateur führt.

Bereits bestehende Modelle mentaler Beanspruchung in der Flugverkehrskontrolle betrachten kognitive Prozesse bisher allerdings eher am Rande, machen also sehr wenige oder stark vereinfachte Annahmen, die der Komplexität des Systems und der Aufgabe nicht gerecht werden. Oder aber sie fokussieren auf den sogenannten en-route-Bereich (Lotsen, die die Koordination von Flugzeugen im oberen Luftraum vornehmen), wobei die Befunde hier als nicht vergleichbar mit dem Aufgabenspektrum und der Komplexität des zu kontrollierenden Systems von Towerlotsen anzusehen sind. Auf dem Flughafen finden einerseits sehr viel mehr Interaktionen zwischen Flugzeugen statt. Andererseits ist die Erkennung und Lösung von Konflikten aufgrund der räumlichen Begrenzung hier ungleich schwerer (es stehen nur 2 Dimensionen zur Lösung von Konflikten zur Verfügung) und zeitkritischer.

Diese Arbeit hat sich daher zum Ziel gesetzt, die Modellbildung und Simulation in Mensch-Maschine-Systemen zu nutzen, um die nötigen Voraussetzung einer verlässlichen Erkennung des Zustandes (insbesondere der mentalen Beanspruchung) von Towerlotsen zu schaffen.

[1] Als Expertensysteme werden Computerprogramme bezeichnet, welche den Menschen bei der Lösung komplexer Probleme auf Grundlage einer vorgegebenen Wissensbasis (Wenn-Dann-Beziehungen) unterstützen.

1.2 Reichweite der Arbeit

Im Kern dieser Arbeit steht die Anwendung der Modellbildung und Simulation zur Beschreibung mentaler Beanspruchung von Towerlotsen. Es gilt vor allem Defizite des vorhandenen Methodenrepertoires zur Erfassung mentaler Beanspruchung zu beseitigen, beispielsweise, dass mit klassischen Methoden keine Vorhersagen über die Beanspruchung gemacht werden können und sie damit keine prospektive Beanspruchungsschätzung zulassen. Die Modellierung ist in der Lage diese Anforderung zu erfüllen (vgl. Kapitel 3.4), da durch Simulationen auch zukünftige Situationen und Bedingungen adressiert werden können. Nicht zuletzt ist der Anspruch ebenso eine Verschmelzung von Methoden ingenieurwissenschaftlicher Anwendungsforschung und Methoden psychologischer und kognitionswissenschaftlicher Grundlagenforschung und damit der Interdisziplinarität der Modellbildung in Mensch-Maschine-Systemen Rechnung zu tragen und von ihr zu profitieren. So wurden die theoretischen Vorarbeiten des Modells unter der Verwendung ingenieurwissenschaftlicher Methoden (Aufgabenanalysen und Feldbeobachtungen; vgl. Kapitel 4.3; 4.4) mit Hilfe eines Beschreibungsmittels aus der Informatik (der farbigen Petrinetze; vgl. Kapitel 4.6) auf der Grundlage kognitionswissenschaftlicher Theorien entwickelt und unter Verwendung psychologischer Methoden zur Versuchsplanung und –auswertung auf Validität geprüft (vgl. Kapitel 6).

In dieser Arbeit wird ein Modell vorgestellt und validiert, welches einen ersten Ansatz darstellt, wie die Modellierung und Simulation kognitiver Prozesse zur Erkennung mentaler Beanspruchung von Towerlotsen angewendet werden kann. Der Nutzen des entwickelten Modells bzw. des vorgestellten Ansatzes kann, seinerseits sowohl anwendungs- als auch grundlagenorientiert, einer Vielzahl Anwendungsmöglichkeiten zugeführt werden. So kann es zur Zustandserkennung von Towerlotsen für die Gestaltung adaptiver Automationen dienen. Darüber hinaus kann es verwendet werden, um neuartige Unterstützungssysteme auf ihren Nutzen hinsichtlich der Entlastung kognitiver Kapazitäten modellhaft zu überprüfen, ohne kostenintensive Prototypenentwicklungen und Nutzertests vornehmen zu müssen. Zusätzlich könnte es Anwendung bei der Flugverkehrsplanung finden, wobei die kognitive Kapazität des Lotsen bei der Erstellung von Flugplänen berücksichtigt werden kann. Ferner kann das Modell ebenso helfen, die kognitiven Vorgänge von Lotsen und von Menschen im Allgemeinen bei der Bearbeitung kognitiver Aufgaben besser verstehen zu können und neue Forschungsfragen zu generieren.

1.3 Forschungsziele

Forschungsziel 1: Ableitung einer geeigneten Modellierungsform des Konstruktes mentaler Beanspruchung aus einschlägigen kognitionswissenschaftlichen Theorien (theoretisch)

Durch eine intensive Recherche der Theorien und Modellvorstellungen zum Konstrukt mentaler Beanspruchung soll eine geeignete Möglichkeit gefunden werden, dieses Konstrukt mit Hilfe der kognitiven Modellierung psychologisch plausibel abzubilden. Hierzu ist eine Informationssammlung einschlägiger Theorien zu mentaler Beanspruchung und den Begrenzungen der menschlichen Informationsverarbeitungskapazität notwendig, um eine adäquate Abbildung mentaler Beanspruchung und menschlicher Informationsverarbeitungs-kapazität in einem kognitiven Modell zu erreichen.

Forschungsziel 2: Verfolgung eines makrokognitiven Modellierungsansatzes (anwendungsorientiert) und Verbindung mit mikrokognitiven Ansätzen (theoretisch)

Es soll die Funktionsweise des Mensch-Maschine-Systems herausgearbeitet werden. Zum Verständnis der Aufgaben und kognitiven Prozessen des Lotsen ist eine intensive Analyse des Systems Flugverkehrskontrolle mit Methoden der Systemtechnik durchzuführen. Gleichzeitig soll durch die Anwendung kognitionspsychologischer Theorien der Informationsverarbeitung und mentaler Beanspruchung auf die Domäne Flugverkehrskontrolle eine Übertragung dieser Theorien auf die Aufgaben und kognitiven Prozesse der Domainexperten erreicht werden. Die Verknüpfung der kognitionspsychologischen Theorien der Informationsverarbeitung mit den zur Aufgabenbearbeitung notwendigen kognitiven Prozessen des Lotsen bildet die Basis für die Implementierung eines kognitiv plausiblen, ganzheitlichen, makrokognitiven Modells.

Forschungsziel 3: Erstellung eines Modells zur quantitativen Beschreibung mentaler Beanspruchung von Towerlotsen in Abhängigkeit verschiedener Einflussfaktoren (anwendungsorientiert)

Auf Grundlage der theoretischen Analysen und Analysen der Anwendungsdomäne soll die Erstellung eines ablauffähigen, ganzheitlichen, kognitiven Modells erfolgen, welches in der Lage ist, sowohl die Anwendungsdomäne Flugverkehrskontrolle, als auch die theoretischen Modellvorstellungen mentaler Beanspruchung und menschlicher Informationsverarbeitung adäquat abzubilden und Vorhersagen zur mentalen Beanspruchung zu ermöglichen. Hierbei soll ein Grundlagenmodell (im Folgenden mit MATriCS; Model of Airport Traffic Control System) basaler Lotsenbeanspruchung und Informationsverarbeitung geschaffen werden, welches die Implementierung weiterer Einflussfaktoren mentaler Beanspruchung ermöglicht.

Forschungsziel 4: Validierung des erstellten Modellansatzes mittels statistischer Methoden (empirisch)

Um zu ermitteln, ob das entwickelte Modell in der Lage ist, valide Vorhersagen der mentalen Beanspruchung von Towerlotsen zu generieren, sollen in einer experimentellen Studie Daten mentaler Beanspruchung erhoben und diese mit den vom Modell generierten Daten verglichen werden.

1.4 Aufbau der Arbeit

Die Arbeit stellt zunächst ausgewählte theoretische Modelle mentaler Beanspruchung und Erfassungsmethoden derselben dar (Kapitel 2). Ferner wird der Stand der Forschung zur komputationalen Modellbildung in der Flugverkehrskontrolle im Allgemein und zur mentalen Beanspruchung im Speziellen vorgestellt (Kapitel 3). Die Vorstellung der theoretischen Vorarbeiten zur Analyse des Systems Flugverkehrskontrolle, zur theoretischen Modellbildung und zur Auswahl des Beschreibungsmittels erfolgt in Kapitel 4. Eine genaue Beschreibung des erstellten Modells, sowie der Studie zu dessen Validierung erfolgen in Kapitel 5 bzw. 6. Kapitel 7 diskutiert die eigene Arbeit, bevor in Kapitel 8 ein abschließender Ausblick auf weitere Forschung gegeben wird.

Kapitel 2 fasst einschlägige kognitionspsychologische Theorien und Modellvorstellungen zum Konstrukt mentaler Beanspruchung zusammen und versucht sich der Beantwortung der Frage zu nähern, was mentale Beanspruchung ist. Hierzu werden verschiedene Modellvorstellungen der psychologischen Grundlagen- aber auch aus der Anwendungsforschung in der Flugverkehrskontrolle erläutert und kontrastiert. Es wird eine Arbeitsdefinition erstellt, welche die Grundlage für eine sowohl psychologisch plausible, als auch anwendungsorientierte Modellierung mentaler Beanspruchung bildet. Ferner werden Methoden zur Erfassung mentaler Beanspruchung vorgestellt und hinsichtlich wichtiger Gütekriterien diskutiert.

Kapitel 3 beschreibt den Stand der Forschung im Bereich Modellierung und Simulation in Mensch-Maschine-Systemen und sensibilisiert für die Notwendigkeit der ganzheitlichen Betrachtung solcher Systeme bei deren Modellierung. Hierbei werden zwei verschiedene theoretische Ansätze zur Modellierung und Simulation in Mensch-Maschine-Systemen vorgestellt, die in der Mensch-Maschine-Forschung und Modellierung häufig kontrastiert werden: mikrokognitive und makrokognitive Ansätze. Es wird argumentiert, dass beide Ansätze zwar verschiedene Wurzeln besitzen, unterschiedliche Zielsetzungen mit der Modellierung verfolgt werden, verschiedene Ansichten bestehen, was ein Modell leisten sollte und daher auch verschiedene Methoden zum Einsatz kommen. Es wird aber auch argumentiert, dass beide Ansätze keinesfalls gegensätzlich sind und damit eine Verbindung beider Ansätze durchaus möglich und auch im Rahmen dieser Arbeit sinnvoll ist. Ferner werden ausgewählte Modelle und Simulationen spezifischer Problemstellungen zur Flugverkehrskontrolle und Luftfahrt, sowie im speziellen zur mentalen Beanspruchung vorgestellt und diskutiert.

Kapitel 4 beschreibt die theoretischen Vorarbeiten zur Erstellung des MATriCS-Modells, die anhand der Richtlinien zur Erstellung kognitiver Simulationen nach Cacciabue (1998) erfolgten. Hierbei wurden zunächst die Problemgrenzen und das Simulationsziel definiert, sowie im Anschluss Aufgabenanalysen der Flughafenverkehrskontrolle zusammengefasst, um den Status-

Quo der Aufgaben von Towerlotsen und der nötigen kognitiven Prozesse zu erfassen. Eine Feldstudie der Arbeitsumgebung wurde durchgeführt, um die Gegebenheiten und Regularien des Systems Flughafenverkehrskontrolle möglichst genau kennen zu lernen. Im Anschluss wurde ein theoretisches Modell erstellt, welches den prototypischen Arbeitsablauf eines Towerlotsen beschreibt und die Einzelhandlungen anhand der Phasen der Handlungsregulation systematisiert und als Grundlage für die Implementierung des Modells dient. Ferner wird die Auswahl des verwendeten Beschreibungsmittels begründet.

Kapitel 5 beschreibt die Implementierung und die Funktionsweise des MATriCS-Modells anhand der genauen Darstellung der Sub-Module auf der Basis farbiger Petrinetze. Hierbei werden die genauen Abläufe des MATriCS-Modells beschrieben, sowie die auf Grundlage verschiedener kognitionspsychologischer Theorien getroffenen Annahmen dargestellt. Ferner erfolgt eine kurze Beschreibung der Funktionsweise der Generierung und des Auslesens von Daten aus dem MATriCS-Modell.

Kapitel 6 stellt die Studie zur Überprüfung der Modellgüte vor. Es wird zunächst untersucht, ob die experimentelle Manipulation erfolgreich war. Die Modellgüte wird auf Grund der Eigenschaft der Daten vor allem durch Trendmaße ermittelt. Hierbei wird einerseits untersucht, ob das Modell in der Lage ist, die Trends der mittleren Beanspruchungshöhe verschiedener subjektiver Maße und des Leistungsmaßes zu replizieren. Ferner wird untersucht, inwieweit das Modell in der Lage ist, die Beanspruchung jeder einzelnen Versuchsperson, erfasst durch ein subjektives Maß und ein physiologisches Maß, adäquat abzubilden. Es folgt eine Diskussion der Ergebnisse.

Kapitel 7 bewertet den erreichten Beitrag der eigenen Arbeit hinsichtlich der Erreichung der selbst gestellten Forschungsziele.

Kapitel 8 gibt einen Ausblick auf weitere Forschungsmöglichkeiten.

2 Mentale Beanspruchung, Kognitive Kapazität und Arbeitsgedächtnis

Das vorliegende Kapitel beschäftigt sich mit der Bedeutung von mentaler Beanspruchung, sowie der Kapazität menschlicher Informationsverarbeitung. Die Betrachtung einschlägiger Theorien ermöglicht einerseits das Aufstellen einer Arbeitsdefinition, andererseits die Ermittlung einer angemessenen Form mentaler Beanspruchung modellhaft abzubilden. In Kapitel 2.1 wird zunächst auf die zwei dominierenden Modellvorstellungen mentaler Beanspruchung eingegangen – *aktivierungstheoretische* und *aufmerksamkeitstheoretische*. Kapitel 2.2 geht im Anschluss detaillierter auf die zentrale Verarbeitung von Informationen im Arbeitsgedächtnis ein, wobei vor allem die Kapazitätsbegrenzungen und die Möglichkeit der Kapazitätserweiterung im Vordergrund stehen. Kapitel 2.3 beschäftigt sich mit der Übertragung der Befunde mentaler Beanspruchung auf den Bereich der Flugverkehrskontrolle. In Kapitel 2.4 werden ausgewählte Methoden zur Erhebung mentaler Beanspruchung vorgestellt sowie hinsichtlich wichtiger Gütekriterien bewertet. Kapitel 2.5 fasst das Kapitel zusammen.

2.1 Mentale Beanspruchung – Theorien und Modelle

In Abgrenzung zu physischer und emotionaler Beanspruchung werden als mental beanspruchend solche Aufgaben beschrieben, die primär durch geistige Anforderungen ein subjektives Gefühl des Angestrengtseins hervorrufen und damit Anforderungen an verschiedene Teile des menschlichen Informationsverarbeitungssystems stellen (Gopher & Donchin, 1986). Häufig wird eine Unterscheidung zwischen Belastung und Beanspruchung dadurch vorgenommen, dass auf bestimmte, von außen auf den Menschen einwirkende Faktoren (sogenannte Belastungs- oder Inputfaktoren; engl. demands, stressors; Manzey, 1998) Bezug genommen wird, welche eine Beanspruchungsreaktion (engl. workload, strain) auslösen. Diese Ansicht schlägt sich auch im Belastungs-Beanspruchungskonzept der Ergonomie nieder (Rohmert, 1984). Als mental beanspruchend wird dabei nur derjenige Teil der Aufgabe bezeichnet, „*der von den aufgabenspezifischen Anforderungen an das menschliche Informationsverarbeitungssystem und damit von der Aufgabenschwierigkeit bzw. –komplexität determiniert wird*“ (Manzey, 1998, S. 800).

Es ist bisher nicht gelungen eine einheitliche und umfassende Theorie mentaler Beanspruchung zu erarbeiten (Manzey, 1998). Vielmehr existieren eine Reihe verschiedener Modellvorstellungen, welche bestimmte Prozesse und Begriffe fokussieren. Stein (1998) reflektiert diesen Sachverhalt etwas ironisch mit: „*…there have been so many workload reviews, that soon we will be reduced to writing reviews of reviews”* (S. 156). Manzey (1998) identifiziert zwei dominierende theoretische Ansätze: Die aus der psychophysiologischen

Aktivierungsforschung stammenden *aktivierungstehoretischen Modelle* (z.B. Boucsein, 1991; Hockey, 1984; Pribram & McGuniess, 1975), welche mentale Beanspruchung als aktivierungstheoretisches Konzept interpretieren (Manzey, 1998), sowie *aufmerksamkeitstheoretische Modelle* (z.B. Kahnemann, 1973; Shiffrin & Schneider, 1977; Wickens, 1984), die aus der Kognitionspsychologie stammend vor allem strukturelle und kognitive Aspekte, wie kapazitätslimitierte kognitive Ressourcen hervorheben (Ribback, 2003).

2.1.1 AKTIVIERUNGSTHEORETISCHE MODELLE

Die psychophysiologische Aktivierungstheorie (z.B. Duffy, 1951; Lindsley, 1951) geht davon aus, dass Informationsverarbeitung Energie benötigt (Ribback, 2003). Diese Energie ist dabei als *„metabolische Kapazität definiert, die zur Aufgabenbearbeitung notwendig ist"* (Ribback, 2003, S. 30). Es wird angenommen, dass mentale Beanspruchung direkt mit dem psychophysischen Aufwand verbunden ist, der für die Bearbeitung komplexer Aufgaben benötigt wird und somit auch direkt aus der Ausprägung psychophysiologischer Aktivierungsindikatoren (z.B. Herzrate, Blutdruck usw.) abzulesen sei (Manzey, 1998; Ribback, 2003). Im Bereich der aktivierungstheoretischen Konzepte konnte im Laufe der Zeit allerdings weder die Vorstellung eines eindimensionalen Kontinuums menschlicher Aktivierung, welches von tiefstem Schlaf bis zum Zustand höchster Erregung reicht, noch ein umgekehrt u-förmiger Zusammenhang zwischen der Leistung und dem Aktivierungsniveau (Yerkes & Dodson, 1908) aufrechterhalten werden. Dies liegt einerseits darin begründet, dass die Übertragung des Yerkes-Dodson Gesetzes auf die menschliche Leistung und Aktivierung fragwürdig erscheint, da Yerkes und Dodson (1908) ihre Untersuchungen an Mäusen durchführten (Dekker & Hollnagel, 2004; Hockey, 1984; Staal, 2004). Andererseits sprechen zahlreiche Befunde dafür, dass in Abhängigkeit von der Auslösebedingung (Stimulus), der individuellen Dispositionen (Individualspezifität) und aktuellen Reaktionstendenzen einer Person (Motivationsspezifität) bestimmte psychophysiologische Reaktionsmuster ausgelöst werden (Fahrenberg, 1983; Manzey, 1998), die in ihrer Ausprägung allerdings unabhängig von der Stärke des Stimulus sind. Kurz gesagt bedeutet dies, dass eine Person auf einen bestimmten Reiz ein gleichartiges Reaktionsmuster zeigt. Diese führten zu einer Erweiterung und Ausdifferenzierung des Aktivierungskonzeptes.

So erweiterten unter Anderem Pribram und McGuniess (1975) diese Konzepte und beschreiben drei verschiedene zentralnervöse Funktionssysteme: Einem *Affect-Arousal-System*, welches hauptsächlich der Steuerung selektiver Aufmerksamkeitsprozesse dienen soll (Manzey, 1998; Ribback, 2003); ein *Preparatory-Activation-System*, welches vor allem die motorische Reaktionsbereitschaft erhöht und ein *Effort-System*, dem zum einen eine richtungssteuernde Funktion der Aufmerksamkeit zukommt (Ribback, 2003). Zusätzlich wird letzterem auch eine koordinierende Funktion zugeschrieben (Manzey, 1998). In diesem Modell wird Anstrengung

durch die wahrgenommene Schwierigkeit der Aufgabe bestimmt. Es wird u.a. durch neurobiologische Befunde untermauert (Robbins & Everitt, 1994), da die Unterscheidung der drei Aktivierungssysteme mit den in der Neurobiologie inzwischen gut identifizierten monoaminergen[2] Bahnsystemen große Ähnlichkeit aufweist.

Stark angelehnt an Pribram und McGuniess (1975) beschreibt Boucsein (1991) drei Systeme, die stark vereinfachend als allgemeine (effort), emotionale (arousal) und motivationale/ tätigkeitsbezogene (activation) Aktivierungsprozesse beschrieben werden können. Diese neueren aktivierungstheoretischen Modelle berücksichtigen den Einfluss intentional-motivationaler Prozesse, in Form von willentlich gesteuerten Kontrollprozessen. Hiermit lässt sich erklären, dass unter mentaler Beanspruchung bestimmte Indikatoren erhöhten Energiebedarfes in der Peripherie (z.B. Blutdruck, Muskelspannung) ansteigen, während im Gehirn, wo die Informationsverarbeitung stattfindet, der Energiebedarf im Vergleich zu einer Ruhebedingung gleich bleibt (Ribback, 2003).

2.1.2 Aufmerksamkeitstheoretische Modelle

Die *aufmerksamkeitstheoretischen* Modelle entstammen der Forschung zu Interferenzeffekten bei der gleichzeitigen Bearbeitung zweier Aufgaben (Manzey, 1998). Hierbei werden Leistungseinbußen bei solchen Doppeltätigkeiten dadurch erklärt, dass das menschliche Informationsverarbeitungssystem kapazitätsbegrenzt ist. Die vorhandene Kapazität reicht bei der Bearbeitung zweier Aufgaben nicht aus, um beide Aufgaben ohne Leistungseinbußen zu bearbeiten (Kramar & Spinks, 1991; Manzey, 1988; Wickens, 1984, 1991). Der Kapazitätsbedarf einer Aufgabe bestimmt dabei das Ausmaß der mentalen Beanspruchung (Manzey, 1998). Je höher dabei der Kapazitätsbedarf, desto geringer ist die jeweilige Restkapazität die noch für weitere Aufgaben zur Verfügung steht (Manzey, 1998; Ribback, 2003).

Dabei unterscheiden sich die Modellannahmen darin, welche Art der Kapazitätsbegrenzung beschrieben wird. So geht beispielsweise Kahnemann (1973) von einer einzelnen, undifferenzierten Kapazität aus, von der Ressourcen für die Aufgabenbearbeitung in begrenztem Umfang zur Verfügung gestellt werden können. Nach Wickens (1992) definiert sich die Kapazität als die maximale Verarbeitungskapazität, während Ressourcen den mentalen Aufwand beschreiben, der für die Verarbeitungseffizienz zur Verfügung gestellt wird (De Waard, 1996). Bei Kahnemann (1973) sind alle Verarbeitungsaufgaben an diese eine Kapazität geknüpft (Ribback, 2003). Sie ist abhängig von dem Ressourcenbedarf, d.h. der Schwierigkeit der Aufgabe, der überdauernden Dispositionen (z.B. nicht willkürlich beeinflussbarer Bereitstellung von Ressourcen bei Orientierungs- oder Defensivreaktionen) und den jeweiligen Intentionen des

[2] Monoaminerge Bahnen umfassen Zellen, welche Noradrenalin, Dopamin, Serotonin oder Histamin als Neurotransmitter verwenden (Robbins, 1997).

Individuums (Manzey, 1998). Bereits Kahnemann (1973) nahm an, dass, obwohl die Gesamtkapazität begrenzt ist, sie bis zu einem gewissen Grad durch Anstrengung gesteigert werden kann. Hierbei nimmt er an, dass das Aktivierungsniveau als Korrelat der verfügbaren Maximalkapazität verstanden werden kann und stellte so bereits einen Zusammenhang zu aktivierungstheoretischen Konzepten her (Manzey, 1998).

Norman und Bobrow (1975) sowie Shiffrin und Schneider (1977) erweiterten diese Theorie und erklären Informationsverarbeitung nun nicht mehr nur mit ressourcentheoretischen Begriffen, sondern sehr viel allgemeiner und fassen darunter auch Anstrengung, Gedächtniskapazität und Kommunikationskanäle. Der hypothetische Zusammenhang zwischen der Ressourcenverteilung und der Leistung in der Aufgabe wird solange als linear angenommen, bis alle Ressourcen investiert sind. Ist dies der Fall, können keine weiteren Ressourcen investiert werden und die Leistung bleibt stabil. Nach Norman und Bobrow (1975) werden solche Aufgaben als ressourcenlimitiert bezeichnet. Diese stehen im Gegensatz zu datenlimitierten Aufgaben (De Waard, 1996). Bei der Ausführung datenlimitierter Aufgaben wird die zusätzliche Investition von Ressourcen keinen Zuwachs in der Leistung bringen, da die Qualität der Daten begrenzt ist (z.B. der aufzunehmenden Informationen). Dieser Zusammenhang zwischen der Leistungsanforderung durch die Aufgabe und kognitiven Ressourcen wird durch die Leistungs-Ressourcen-Funktion (performance-ressource-function, PRF) beschrieben, die in einen ressourcenlimitierten und einen datenlimitierten Teil unterteilt wird. Interferenzeffekte bei der gleichzeitigen Bearbeitung von zwei Aufgaben interpretieren Norman und Bobrow (1975) dahingehend, dass ressourcenlimitierte Verarbeitungsprozesse mit anderen kognitiven Prozessen konkurrieren. Nichtsdestotrotz bleibt mentale Beanspruchung nach diesen Theorien ein eindimensionales Konzept (Mulder, 1980).

Obwohl die bisher beschriebenen Theorien auf eine Vielzahl von Situationen übertragen werden können, können sie nicht erklären, warum unbeeinflusste Leistung und das gleichzeitige Ausführen mehrerer Aufgaben möglich sind (De Waard, 1996). Zahlreiche Experimente zu Doppeltätigkeiten widersprachen der Auffassung eines undifferenzierten Kapazitätskonzeptes, wie es Kahnemann (1973) postulierte und dem auch Norman und Bobrow (1975) folgten. Es konnte gezeigt werden, dass Interferenzeffekte bei der gleichzeitigen Ausführung zweier Aufgaben durch die Ähnlichkeit in der Anforderungsstruktur der Aufgaben bestimmt werden, wobei die Aufgabenschwierigkeit eine eher untergeordnete Rolle spielt (Heuer, 1996; Manzey, 1988; Wickens, 1980, 1984). Zwei Aufgaben interferieren demnach nur dann, wenn sie ähnliche Anforderungen an den kognitiven Apparat stellen. Nur dann wirkt sich auch ihre Schwierigkeit auf die Leistung aus (Manzey, 1988).

Dies führte zur Entwicklung der mehrdimensionalen Ressourcentheorien (z.B. Navon & Gopher, 1979; Ploson & Friedman, 1988; Wickens, 1984). Hierbei wird davon ausgegangen, dass

das menschliche Informationsverarbeitungssystem über mehrere, voneinander unabhängige Teilsysteme verfügt, welche jeweils für sich über begrenzte Ressourcen verfügen. Interferenzen bei Doppelaufgaben entstehen dadurch, dass zwei strukturell sehr ähnliche Aufgaben gleichzeitig bearbeitet werden und damit ähnliche Verarbeitungsressourcen benötigen.

Wickens (1984, 1991, 2002; vgl. auch Wickens & McCarley, 2008) postuliert ein vierdimensionales multiples Ressourcenmodell, wobei jede der vier Dimensionen weiterhin zwei diskrete Stufen aufweist. Die erste Dimension stellen hierbei die beiden Stufen der Informationsverarbeitung und deren Begrenzungen dar: perzeptiv-kognitive Prozesse, also die Informationsaufnahme (z.B. visuelles Scanning; Manzey, 1998) und die zentrale Verarbeitung (insbesondere das Arbeitsgedächtnis; Wickens & McCarley, 2008) sowie die Steuerung der Vorbereitung und Kontrolle motorischer Reaktionen (Handlungsausführung). Als zweite Dimension beschreiben Wickens und Mccarley (2008) die Wahrnehmungsmodalitäten, mit den Stufen auditiv und visuell. Die dritte Dimension beschreibt die beiden Stufen der visuellen Kanäle focal und ambient (Wickens & McCarley, 2008). Als vierte Dimension werden die Verarbeitungscodes mit ihren Stufen, räumlich-analog und kategorisch-symbolisch (linguistisch oder verbal; Wickens & Mccarley, 2008) beschrieben. Es ist dabei darauf hinzuweisen, dass die Unterscheidung der Wahrnehmungsmodalitäten (auditiv; visuell) nur auf dem Level der Perzeption getroffen werden kann, nicht aber auf dem Level der zentralen Verarbeitung. Ebenso wird die Unterscheidung zwischen ambient und focal nur für den visuellen Kanal getätigt. Dies bedeutet, dass für die zentrale Verarbeitung nur die Verarbeitungscodes (räumlich-analog und kategorisch-symbolisch) weiterhin unterschieden werden können und Informationen bei ihrer Verarbeitung damit auch abhängig von ihrer Codierung unterschiedliche Ressourcen beanspruchen.

Ebenfalls vom Modell von Pribram und McGuniess (1975) ausgehend stellt Hockey (1997) ein Modell vor, welches beide theoretischen Ansätze, aktivierungs- und aufmerksamkeits-theoretische, verbindet. Er geht von begrenzten multiplen Ressourcen aus, wobei Anstrengung für eine Energiebereitstellung sorgt (Ribback, 2003). Hockey (1997) nimmt dabei eine Integration der beiden vorgestellten Ansätze vor, woraus sogenannte integrative Theorien folgen, die die begrenzten Verarbeitungsressourcen der ressourcentheoretischen Modelle als energetische Begriffe auffassen und mentale Beanspruchung somit auch der physiologischen Erfassung zugänglich machen (Hockey, 1997; Sanders, 1983; Wickens, 1991).

Aus dem Modell von Wickens (1984) folgt, dass vor allem Begrenzungen der perzeptiv-kognitiven Verarbeitungsstufe für das Entstehen mentaler Beanspruchung verantwortlich sind. Hierbei wird neben der Kapazitätsbegrenzung der informationsaufnehmenden Sinne vor allem das Arbeitsgedächtnis als limitierender Faktor gesehen. Es werden daher im folgenden Kapitel

Modellvorstellung über den Aufbau und die Begrenzungen der Arbeitsgedächtniskapazität näher vorgestellt.

2.2 Arbeitsgedächtnismodelle, Kapazität und Chunks

Grundlegend wird das Arbeitsgedächtnis als Kurzzeitspeicher und Instanz für die Manipulation von Informationen, die zur Bearbeitung von Aufgaben notwendig sind, angesehen. Dieses ist abzugrenzen von sensorischen Speichern der Sinne und dem Langzeitgedächtnis (Sanders & McCormick, 1993). Im sensorischen Speicher werden Eingaben der Sinnesorgane für sehr kurze Zeit aufrechterhalten (wenige Sekunden). Wird Aufmerksamkeit darauf gelegt, werden diese Informationen in das Arbeitsgedächtnis überführt. Vom Arbeitsgedächtnis können Informationen in das Langzeitgedächtnis durch semantische Kodierung bzw. Sinnanwendung und durch Herstellung von Beziehungen zu bereits existierendem Wissen übertragen werden (Vingelis, Schaeffer, Stringer, Gromelski, & Ahmed, 1990).

Die Vorstellung multipler Ressourcen, die auch Wickens (1984, 1991) postuliert, spiegelt sich ebenso in vielen Modellvorstellungen über den Aufbau und die Funktion der zentralen Verarbeitung von Informationen wieder. Auch hierbei deuten neuere Forschungsergebnisse darauf hin, dass die Speicherung und Verarbeitung von Informationen nicht nur von einer einzelnen limitierenden Ressource mit bestimmter Kapazität abhängt, sondern, dass vielmehr mehrere unabhängig voneinander agierende Systeme mit jeweils eigener Kapazität agieren.

Baddeley und Hitch (1974) beschreiben dabei ein Arbeitsgedächtnismodell, welches über eine *zentrale Exekutive* verfügt, die zwei untergeordnete Systeme steuert: die *phonologische Schleife* und den *visuell-räumlichen Notizblock* (Baddeley, Grant, Wight, & Thomson, 1975; Logie, 1986). Dieses Modell wurde durch die Gewinnung neuer Erkenntnisse über die Jahre hinweg immer wieder angepasst. So wurde später noch ein drittes System, der *episodische Puffer*, hinzugefügt (Baddeley, 2000). Baddeleys Modell ist dabei ebenso ein multiples Ressourcen Modell wie das von Wickens (Stein, Garland, & Muller, 2010).

Bei Baddeley liegt ebenso wie bei Wickens (1984) die Annahme zugrunde, dass die jeweiligen Komponenten des Arbeitsgedächtnisses unabhängige Ressourcen beanspruchen, welche ihrerseits in gewisser Weise limitiert sind. Beide machen aber keine genaue Aussage darüber, wie groß diese Ressourcenreservoirs sind. Für eine angestrebte modelhafte Beschreibung mentaler Beanspruchung und kognitiver Kapazität wäre es nicht von unerheblichem Interesse zu analysieren, wie hoch die maximale für die Aufgabenbearbeitung zur Verfügung stehende Kapazität eines Menschen ist. Dies ist allerdings eine äußerst schwierige Aufgabe. Beispielsweise gibt Wickens (1984) keinerlei Hinweise darauf, wie groß die jeweiligen Ressourcenreservoirs der Perzeption, sowie der zentralen Verarbeitung sein könnten. Für die Perzeption kann

allerdings von rein physikalischen Begrenzungen ausgegangen werden. Auch wenn es trivial erscheinen mag, wir haben nur ein Paar Augen und Ohren und können unsere fokussierte Aufmerksamkeit (focal) nur auf einen Sachverhalt gleichzeitig lenken (vgl. z.B. Posner, Snyder, & Davidson, 1980). Ebenso kann nur einer Person zugehört werden, ohne dass durch Interferenzeffekte Informationen verloren gehen (vgl. z.B. Baddeley, 1976).

Etwas anders ist die Sachlage bei der Untersuchung der zentralen Verarbeitung von Informationen, speziell denen des Arbeitsgedächtnisses. Hier wurden eine Reihe von Untersuchungen durchgeführt und Theorien aufgestellt, wie groß das maximale Ressourcenreservoir der Informationsverarbeitung ist. Diese werden im Folgenden beschrieben.

Bei der zentralen Verarbeitung von Informationen ist es sehr naheliegend, dass mehrere Informationseinheiten über einen gewissen Zeitraum aufrechterhalten werden müssen, um sie effektiv manipulieren zu können. So müssen beispielsweise für die Multiplikation beide Faktoren aufrechterhalten werden, um das Produkt zu bilden.

Im oben beschriebenen Modell von Baddeley (Baddeley & Hitch, 1974; Baddeley, 2012) werden nur sehr wenige Hinweise darauf gegeben, ob eine Ressourcenobergrenze existiert und wenn ja, wie groß diese ist. Bei Baddeley wird lediglich dem episodischen Puffer eine begrenzte Kapazität von vier Informationseinheiten (sogenannten Chunks;) zugesprochen. Die Arbeit von Miller (1956) hat hierbei maßgeblich die Entwicklung des Konzepts der Zusammenfassung einzelner Informationseinheiten zu einem Gefüge, genannt „Chunks“ bestimmt, sowie außerdem den richtungsweisenden Anstoß für den information-processing approach gegeben (Baddeley, 1994). Der Begriff des Chunk wird bis heute verwendet, um die Zusammenfassung mehrerer Informationseinheiten (Items; z.B. Ziffern, Buchstaben) zu beschreiben, die in einen solchen Sinnzusammenhang gebracht werden, dass sie als ein einzelnes Element im Arbeitsgedächtnis gespeichert werden können (Stein et al., 2010). „*Hence a familiar and meaningful date such as 1492 is likely to act as one chunk, while a less significant one like 1386 would probably not.*“ (Baddeley, 1990, S. 42). Dieser steht in unmittelbarem Zusammenhang mit der Frage, wie groß die maximale Kapazität des Arbeitsgedächtnisses ist und bietet eine Erklärung dazu, wie diese Kapazität erweitert werden kann. Allerdings ist bis heute nicht ganz klar, wie ein solcher Chunk genau aufgebaut ist und wie groß dieser sein kann (Simon, 1974).

Der oben angesprochenen Annahme von Baddeley (1994), der vier speicherbare Informationseinheiten beschreibt, ist auch bei Cowan (1999) wiederzufinden. Cowan (2001) beschreibt allerdings keinen episodischen Puffer. Vielmehr geht er davon aus, dass ein sogenannter „focus of attention“ existiert, welcher für die Aufrechterhaltung und den Abruf von Informationen zuständig ist. Informationen, die nach einer bestimmten Zeit nicht in diesen Aufmerksamkeitsfokus geraten, verblassen (decay). Die Kapazitätsgrenze liegt hierbei im

Aufmerksamkeitsfokus (diese kann in Baddeleys Modell in der zentralen Exekutive gesehen werden), welcher nur in der Lage ist, maximal vier Elemente gleichzeitig abzurufen (Cowan, 2001, 2010; vgl. auch Broadbent, 1975). Eine solches zeitliches Verblassen von Informationen beschreibt auch Baddeley (2012) für seine anderen Systeme.

Aufbauend auf den Ansätzen von Cowan (2001) beschreiben auch Oberauer, Süß, Schulze, Wilhelm und Wittmann (2000) ein aktives Arbeitsgedächtnis, bei dem auch die Vorstellung eines „focus of attention" verfolgt wird, welcher notwendig ist, um Informationen aktiv zu halten und für die Manipulation zugänglich zu machen (vgl. auch Oberauer, Süß, Wilhelm, & Wittman, 2003; van der Linden, 1998). Es wird allerdings bezweifelt, dass dieser Fokus mehr als eine Informationseinheit gleichzeitig beinhalten kann (Oberauer & Bialkova, 2009; vgl. auch McElree, 2001).

Hinsichtlich der Festlegung einer Obergrenze der Verfügbaren Arbeitsgedächtnisressourcen ist zu erkennen, dass die vorgestellten Multi-Komponenten-Systeme eher vage bzw. uneinig sind. Sie treffen zwar detaillierte Annahmen darüber, wie das Arbeitsgedächtnis aufgebaut ist und welche unabhängigen Sub-Systeme existieren. Es gibt allerdings keine oder zumindest keine einheitlichen Befunde darüber, wie hoch die maximale Kapazität dieser Sub-Systeme ist bzw. wie lange es dauert, bis eine Information zerfällt, wenn sie nicht in den Aufmerksamkeitsfokus gerät.

Frühe Ansätze zur Kapazität des Arbeitsgedächtnisses verfolgen eine sehr viel einfachere Vorstellung und treffen sehr viel konkretere Annahmen über die maximale Kapazität. Während die Modelle von Baddeley (1986), Cowan (1999) und Oberauer und Bialkova (2009) alle von Multi-Komponenten-Systemen ausgehen, beschreiben eine Reihe von Autoren bis heute eine „single, limited-capacity system theory of working memory" (Stein et al., 2010) (z.B. Adams, 1989; Kantowitz & Sorkir, 1983; Klapp, Marshburn, & Lester, 1983; Sanders & McCormick, 1993). Ähnlich wie bei Kahnemann (1973) wird hier von einem einzigen, kapazitätsbegrenzten Ressourcenreservoir ausgegangen, welches das menschliche Arbeitsgedächtnis darstellt. Hierbei wird allerdings ein statisches Arbeitsgedächtnis vorausgesetzt, welches lediglich der Speicherung, nicht aber der Manipulation von Informationseinheiten dient.

Die Arbeit von Miller (1956) wurde hierbei unter Anderem auch dadurch berühmt, dass sie eine der ersten darstellt, die eine Obergrenze an Chunks postuliert, die im Arbeitsgedächtnis gespeichert werden können. Miller (1956) postuliert, dass eine Anzahl von 7±2 Elementen gleichzeitig im Arbeitsgedächtnis speicherbar ist („The magical number seven plus or minus two"). Die „single-system" Theorien gehen weiterhin davon aus, dass keine weiteren Gedächtnisaufgaben ohne Leistungseinbußen erfüllt werden können, wenn das Arbeitsgedächtnis einmal mit seinen fünf bis neun Chunks gefüllt ist. Es stellt daher einen

passiven Kurzzeitspeicher für Informationen dar (Oberauer et al., 2000). Auch Baddeley (1990) spricht von einer „normalen“ Anzahl von sechs oder sieben Chunks. Durch das Bilden sinnvoller Chunks ist allerdings dennoch eine Erhöhung der Anzahl an einzelnen Items, die im Gedächtnis gespeichert werden können, erreichbar. Hierzu muss lediglich die Anzahl der Items in einem Chunk erhöht werden. So konnte gezeigt werden, dass die Gedächtnisspanne für zufällig ausgewählte Buchstaben bei etwa sechs liegt. Diese Anzahl kann auf etwa neun erhöht werden, wenn die Buchstaben in Silben angeordnet sind, die nach dem Schema Konsonant-Vokal-Konsonant aufgebaut sind. Wenn die Buchstaben allerdings Wörter in Sätzen mit semantischer Bedeutung ausmachen, können bis zu 50 oder mehr Buchstaben erinnert werden (Baddeley, 1990). In jedem Fall bleibt aber die Anzahl der Chunks konstant bei etwa sechs (Baddeley, 1990).

Es kann angenommen werden, dass die Fähigkeit, Chunks aus ursprünglich nicht zusammenhängenden Items zu konstruieren, stark vom Wissen, der Erfahrung und dem Training innerhalb einer Aufgabe abhängt (Baddeley, 1990; siehe auch Ericsson & Chase, 1982). Er ist allerdings unabhängig vom Material (z.B. verbal oder räumlich), aus dem er sich zusammensetzt (Simon, 1974; Smieszek, Manske, Hasselberg, Russwinkel, & Möhlenbrink, 2013). Die Fähigkeit, sinnvolle Chunks zu bilden, kann daher bei Spezialisten, für eine spezifische Aufgabe als sehr stark ausgeprägt angenommen werden. Neisser (1979) beschreibt einen solchen Effekt am Beispiel eines Schachmeisters: *„Ein Meister kann beispielsweise die Stellung aller Figuren auf dem Brett angeben, wenn er fünf Sekunden darauf geschaut hat; kein Amateurspieler kann das annähernd gleich gut. Der Meister kommt zum Erfolg, weil er bestimmte Aspekte der Position wahrnehmen kann, die dem schlechten Spieler entgehen; strukturelle Merkmale, die, hat man sie einmal gesehen, den Ort der einzelnen Figuren zwingend bestimmen.“* (Neisser, 1979, S. 141). Eine solche Beobachtung, wie Neisser (1979) sie hier beschreibt, kann damit erklärt werden, dass der Schachmeister bestimmte Konstellationen von Figuren als Chunks zusammenfassen kann und sich nicht, wie der Novize, die Position jeder einzelnen Figur merken muss. Tatsächlich konnten de Groot (1965) sowie Chase und Simon (1973) diese Beobachtung in Experimenten bestätigen.

Studien, die sich mit der maximalen Arbeitsgedächtniskapazität von Lotsen im Bereich Flugverkehrskontrolle beschäftigen, deuten auf eine ähnliche Obergrenze der speicherbaren Chunks im Arbeitsgedächtnis, sowie eine Verbesserte Chunking-Fähigkeit mit wachsender Erfahrung hin (Bisseret, 1971; Sperandio, 1969; Vingelis et al., 1990). In einer Studie bat Sperandio (1969) Towerlotsen im Anschluss an reale Kontrollsessions die Flugzeuge zu wiederholen, die sie in der vergangenen Stunde kontrolliert hatten. Hier zeigte sich, dass im Mittel zehn Flugzeuge wiedergegeben werden konnten. Bisseret (1970, 1971) untersuchte drei Gruppen von Radarlotsen mit unterschiedlicher Berufserfahrung: hochgradig erfahrene Lotsen;

Lotsen, die gerade den niedrigsten Qualifikationstest bestanden hatten; und Trainees mit drei bis sechs Monaten weniger Erfahrung (Bainbridge, 1975). Er konnte zeigen, dass die Lotsen mit sehr viel Erfahrung sich im Mittel an zehn Flugzeuge erinnern konnten, was mit dem Befund von Sperandio (1969) übereinstimmt. Weniger erfahrene Lotsen konnten noch acht Flugzeuge wiedergeben, Trainees sieben. Er konnte weiterhin zeigen, dass bis zu drei Items[3] pro Flugzeug erinnert werden. Weiterhin berichtet Bisseret (1970, 1971), dass die erfahrenen Lotsen insgesamt 33 Items (also ca. 3 Items x 10 Flugzeuge), die qualifizierten Lotsen 23 Items und die Trainees noch 20 Items erinnern konnten. Durch ihre Erfahrung sind Lotsen verstärkt in der Lage, aus den einzelnen Informationseinheiten der Flugzeuge (z.B. Rufzeichen, Ankunfts-/Abflugzeit, Flugzeugtyp) Chunks zu bilden, die repräsentativ für ein Flugzeug stehen (Vingelis et al., 1990). Aus diesen Ergebnissen ist ebenfalls zu vermuten, dass der Expertisegrad nicht nur die Fähigkeit Chunks zu bilden verbessert, sondern auch eine allgemeine Kapazitätserweiterung nach sich zieht. So kann angenommen werden, dass erfahrenere Lotsen einerseits mehr Informationen in einem Chunk zusammenfassen, andererseits aber auch mehr Chunks insgesamt abspeichern können. Hierbei ist für Experten eine mittlere Anzahl von zehn Chunks maximal im Arbeitsgedächtnis speicherbar (vgl. Bainbridge, 1975; Bisseret, 1970, 1971; Sperandio, 1969).

Studien mit en-route-Lotsen zeigten, dass Chunks aus Informationen über Flugzeuge aufgebaut sind, für die ein Kollisionspotential besteht (Redding & Seamster, 1994). Nach Wickens (1999) behandeln Lotsen Flugzeuge als einzelne Einheiten. Wenn allerdings die Anzahl der Flugzeuge steigt, werden Gruppen von Flugzeugen als Einheit behandelt, ähnlich wie beim obigen Schachbeispiel. Ob auch Towerlotsen solche Strategien anwenden gilt es zu untersuchen. Für die vorliegende Arbeit wird angenommen, dass Informationen über ein Flugzeug als ein Chunk abgelegt wird.

2.3 Mentale Beanspruchung in der Flugverkehrskontrolle

Wie bereits einleitend beschrieben, kann das Konzept der mentalen Beanspruchung allerdings nicht nur durch eine Begrenzung der kognitiven Ressourcen des Informationsverarbeitenden Systems Mensch beschrieben werden (Gopher & Donchin, 1986), da die Leistung in der Aufgabe nicht ausschließlich von den maximal zur Verfügung stehenden Ressourcen abhängt. Nach dem bereits erwähnten Belastungs-Beanspruchungs-Konzept (Rohmert, 1984) spielt auch die Aufgabenschwierigkeit oder *Komplexität* eine wichtige Rolle. Diese kommt vor allem beim Auftreten von Interferenzen zwischen verschiedenen Teilaufgaben zum Tragen. Hiervon kann

[3] Hierbei sei nochmals darauf hingewiesen, dass es sich um Items, nicht um Chunks handelt (Items sind hier beispielsweise Eigenschaften von Flugzeugen, wie Position, Flughöhe usw.). Es ist weiterhin davon auszugehen, dass die Anzahl der erinnerten Items dadurch erhöht wird, da sie besser zu sinnvollen Chunks zusammengefasst werden können.

bei Aufgaben in der Flugverkehrskontrolle meist ausgegangen werden. Es handelt sich um ein hoch komplexes System, in dem die gleichzeitige Ausführung mehrerer Aufgaben üblich ist. Lange wurde die mentale Beanspruchung dabei als: *„…a demand placed upon humans."* (DeWaard, 1996, p. 15), also eine Anforderung an den Menschen, angesehen und damit als lediglich von äußeren Faktoren abhängig interpretiert. Daher wurde versucht, aus *einzelnen, objektiv messbaren äußeren Faktoren* (z.B. die Anzahl der zu kontrollierenden Flugzeuge) direkt auf die mentale Beanspruchung von Lotsen schließen zu können (z.B. Athènes, Averty, Puechmorel, Delahaye, & Collet, 2002; Pawlak, Brinton, Crouch, & Lancaster, 1996). Forschung hierzu wurde hauptsächlich im en-route Bereich betrieben. Dabei fokussieren einige Verfahren auf die physische Beanspruchung in Abhängigkeit der Komplexität (Schmidt, 1976; Soede, Coeterier, & Stassen, 1971). Andere beschäftigen sich mit der Messung der Zeit, die Lotsen mit der Kommunikation verbringen als eine Größe für mentale Beanspruchung. Thonrnhill (1995 nach Pawlak et al., 1996) verwendete die Anzahl an Einträgen, die ein Lotse machen muss, die Zeit, die er mit Kommunikation verbringt und andere Verkehrsfaktoren, um eine Messgröße mentaler Belastung und Beanspruchung zu schaffen (Manning, Fox, & Pfleiderer, 2003; Manning, Mills, Fox, Pfleiderer, & Mogilka, 2001). Lee, Feron und Pritchett (2007) definieren die Komplexität anhand der nötigen Kontrollaktivität, welche hier über die Gesamtzahl der Flugrichtungsänderungen gemessen wird. In den meisten Studien, die den Zusammenhang zwischen Belastung und Beanspruchung untersuchen, wurde vor allem die Verkehrsmenge als Belastungsindikator herangezogen (Costa, 1993; Hagemann, 2000).

Häufig gehen Studien aber, wie auch oben schon beschrieben, davon aus, dass die Schwierigkeit der gesamten Aufgaben bzw. die Gesamtkomplexität der Verkehrssituation maßgeblich die mentale Beanspruchung bestimmt, also eine *Vielzahl von Faktoren* zu berücksichtigen sind, nicht nur ein einzelner Faktor. Hierbei wird nicht die mentale Beanspruchung, sondern die Aufgabenkomplexität als von vielerlei äußeren Faktoren abhängige Variable angesehen. Variablen, wie die Anzahl der zu kontrollierenden Flugzeuge oder die Sektorstruktur, beeinflussen hierbei die Komplexität, aus der sich wiederum die mentale Beanspruchung ergibt. Zunächst diente auch hier die schlichte Zählung z.B. der Anzahl an Flugzeugen oder der Anzahl Ankünfte und Abflüge innerhalb eines bestimmten Zeitabschnittes als Maß für die Komplexität einer Verkehrssituation. Es wurde das neue Konstrukt der ATC-Complexity (Flugverkehrskontroll-Komplexität) eingeführt, um ein „altes" Konstrukt, mentale Beanspruchung, zu beschreiben. Es ist allerdings ebenso schwierig Komplexität zu definieren, wie es schwierig ist mentale Beanspruchung zu definieren. Mogford, Guttman, Morrow und Kopardekar (1995) definieren ATC-Complexity als *„construct that is composed of a number of sector and traffic complexity dimensions or factors"* (S. 3). Dabei kann es sich sowohl um physische Aspekte (z.B. Größe und Konfiguration der Flugrouten) als auch Faktoren, die mit der

Bewegung von Flugzeugen durch den Luftraum verbunden sind (z.B. die Zahl steigender und sinkender Flugzeuge; Mogford et al., 1995), handeln. Hinzu kommt die Schwierigkeit, dass auch die Arbeiten, die sich mit dem Komplexitätskonstrukt beschäftigen, keine einheitliche Definition mentaler Beanspruchung voraussetzen (e.g. Athènes et al., 2002; Mogford et al., 1995). Hauptsächlich werden allerdings Definitionen verwendet, die sich an aufmerksamkeits-theoretischen Modellen orientieren. In diesem Zusammenhang beschreiben Athènes et al. (2002) die Entwicklung eines sogenannten Traffic Load Index (TLI), mit dessen Hilfe sie bestimmte Komplexitätsfaktoren gewichten, um ein Maß für die mentale Beanspruchung von Radarlotsen in Abhängigkeit von der Komplexität einer Verkehrssituation zu gewinnen. Für die Berechnung des TLI fließt zunächst jedes Flugzeug, welches sich im zu kontrollierenden Sektor befindet mit einem Wert von 1 ein, um die kognitiven Ressourcen zu repräsentieren, die benötigt werden, um eine Entscheidung zu fällen, was zu tun ist (Athènes et al., 2002). Dieser Wert kann erhöht werden, wenn beispielsweise eine Konfliktlösung erforderlich ist. Daher wird jedes Flugzeug mit einem Wert zwischen 1 (ausschließlich beobachtetes Flugzeug) und 3,5 (Flugzeug im Konflikt) gewertet (Athènes et al., 2002). Dieser Wert wird für jedes Flugzeug und jedes Radarupdate neu berechnet. Solche Ansätze fokussieren allerdings vor allem auf den Einfluss äußerer Faktoren und sind darüber hinaus wenig kognitiv begründet.

Koros, Della Rocco, Panjwani, Ingurgio und D'Arcy (2003, 2006) versuchen solche Ansätze im en-route Bereich auf Towerlotsen zu übertragen. Koros et al. (2003) identifizieren dazu 29 Komplexitätsfaktoren in neun Kategorien und lassen diese von erfahrenen Towerlotsen in ihrer Wichtigkeit einschätzen. Modellhaft beschreiben Koros et al. (2006) den Einfluss von äußeren Faktoren (Ausgangsfaktoren; vor allem Verkehrs- und Flughafencharakteristika) und mediirender Faktoren, wie kognitive Strategien und individuelle Unterschiede, auf die mentale Beanspruchung (siehe Abbildung 1).

Abbildung 1: Kognitives Model der Fluglotsenaufgabe (eigene Darstellung nach Koros el al. 2006; Mogford et al. 1995)

Hierbei ist zu erkennen, dass nicht nur die äußeren Faktoren für die Ausprägung und das Empfinden von mentaler Beanspruchung verantwortlich sind, sondern auch die, bei Koros et al. (2006) als mediierende Faktoren bezeichneten Variablen Berücksichtigung finden müssen. Es spielen dabei die schon genannten individuellen Unterschiede, wie beispielsweise die Erfahrung des Lotsen, eine wichtige Rolle. Diese steht wiederum direkt mit der Fähigkeit sinnvolle Chunks zu bilden und damit den maximal zur Verfügung stehenden kognitiven Ressourcen zur Aufgabenerfüllung in Zusammenhang.

Wie auch im bereits beschriebenen Belastungs-Beanspruchungs-Konzept (Rohmert, 1984) wird auch hier zwischen den von außen auf den Menschen einwirkenden Einflüssen, als *Belastung* bezeichnet, und den individuellen, unmittelbaren und langfristigen, als *Beanspruchung* bezeichneten, Auswirkungen auf den Menschen unterschieden (Hagemann, 2000; vgl. auch Rouse, Edwards, & Hammer, 1993). Hierbei können allerdings nicht nur individuelle Fähigkeiten, sondern beispielsweise Faktoren wie Motivation, angewendete Strategien, sowie Stimmung des Operateurs einen Einfluss auf die Beanspruchung besitzen (De Waard, 1996). Die Belastung wird vom Ziel der Aufgabe bestimmt, ist extern und unabhängig vom Individuum. Beanspruchung hingegen, wird verwendet, um die Effekte zu beschreiben, die durch die Belastung an den Operateur gestellt wird (vgl. Kap 2.1).

Das schlichte Ersetzen des einen Konstruktes (mentale Beanspruchung) durch das Einführen eines weiteren Konstruktes (ATC-Complexity) zwischen den äußeren Faktoren und der im Fokus des Interesses stehenden mentalen Beanspruchung erscheint allerdings wenig zielführend, bedenkt man, dass ATC-Complexity ebenfalls als subjektive Größe interpretiert wird, die ebenso wie mentale Beanspruchung von inneren Faktoren des Lotsen abhängt (Mogford et al., 1995). Gerade diese Charakteristik ist für deren Untersuchung problematisch: Eine bestimmte Aufgabe kann für einen Menschen mit seinen individuellen Erfahrungen, seinem Wissen und Fähigkeiten wenig beanspruchend sein kann, während sie für einen anderen Menschen mit anderen Erfahrungen, anderem Wissen und anderen Fähigkeiten sehr beanspruchend sein kann.

Ein weiteres Problem, welches sich bei genauer Betrachtung des Modells von Koros et al. (2006) ergibt, ist die sehr vereinfachte Vorstellung, die es repräsentiert. Mentale Beanspruchung wird als Resultat der Eingangsvariablen in Zusammenwirkung mit den mediierenden Faktoren beschrieben. Dies ist insofern korrekt, als dass natürlich die Ausprägung bestimmter „äußerer" und „innerer" Variablen die empfundene, mentale Beanspruchung eines Menschen beeinflusst. Die Abbildung 1 impliziert aber auch, dass dieses Resultat der mentalen Beanspruchung einer direkten Messung zugänglich wäre. Gopher & Donchin (1986) beispielsweis weisen aber ausdrücklich darauf hin, dass sich mentale Beanspruchung auf *„processes or entities that are not directly observable."* (S. 41-4) bezieht. Schmidtke (2002) berichtet weiterhin, dass es noch keine

gesicherten Erkenntnisse *„für mentale Belastung und Beanspruchung durch Aufnahme, Verarbeitung und Umsetzung von Informationen gibt“* (S. 5), dass also keine Erkenntnisse darüber existieren, wie sich eine bestimmte Belastung in einer bestimmten Beanspruchung manifestiert und wie diese Prozesse mental ablaufen. Solche gesicherten Erkenntnisse lassen sich laut Schmidtke (2002) nur dann gewinnen, wenn der *„Forschungsgegenstand einer wie auch immer gearteten Messung zugänglich ist“* (S. 5). Dies ist beim Konstrukt mentale Beanspruchung allerdings nicht der Fall. Es besteht lediglich die Möglichkeit sie indirekt über bestimmte Verfahren erfassbar zu machen, welche als hinreichend diagnostisch erachtet werden, um Rückschlüsse darauf zu erlauben, wie hoch die mentale Beanspruchung eines Menschen in einer bestimmten Aufgabe war. Auf Grund dieser Eigenschaft des Konstruktes, kritisieren u.a. Hollnagel und Woods (2005) dessen Verwendung stark. Sie bezeichnen es als „folk model“, welches zwar einen wichtigen Aspekt des Zustandes von Menschen beschreibt, sich aber auf sogenannte „intervenierende“ Variablen bezieht (vgl. Dekker & Hollnagel, 2004; Dekker & Woods, 2002). Es wird versucht „dazwischenliegende“ mentale Zustände zu erfassen, nicht aber die beobachtbare Performanz (Hollnagel & Woods, 2005; vgl. Abbildung 2).

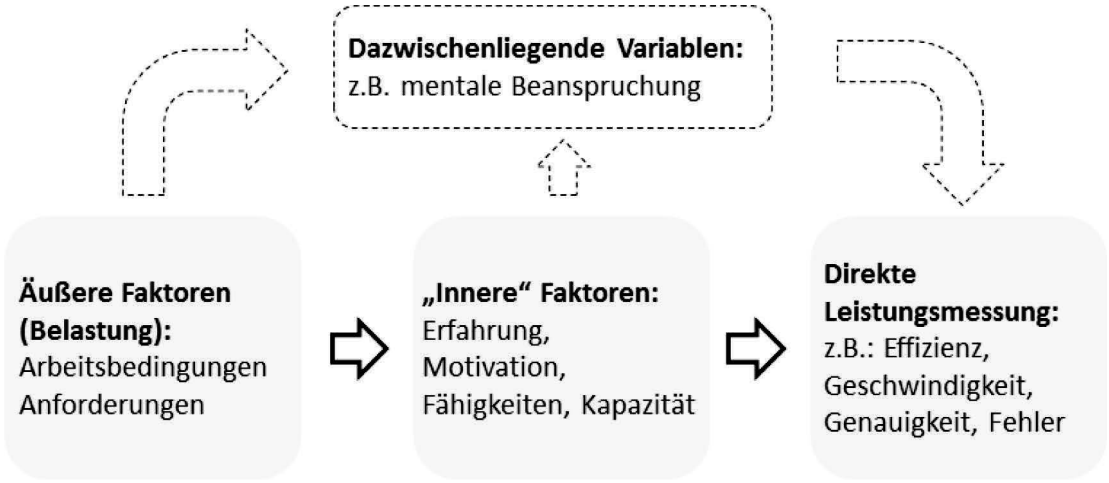

Abbildung 2: Zusammenhang zwischen Eingangsfaktoren, mentaler Beanspruchung und direkt messbaren Variablen (eigene Darstellung nach Hollnagel & Woods, 2005)

Wie in Abbildung 2 zu erkennen, lässt nur eine direkte Leistungsmessung einen Rückschluss auf diese „dazwischenliegenden“ Variablen zu. Es kann also nur von einer Erfassung mentaler Beanspruchung im weitesten Sinne die Rede sein, nicht aber von einer Messung im physikalischen Sinne. Soll die mentale Beanspruchung einer solchen Erfassung zugeführt werden, so muss zumindest eine Arbeitshypothese darüber existieren, wie der menschliche Organismus auf eine von außen einwirkende Belastung reagiert (Schmidtke, 2002). Dies wurde bereits innerhalb dieses Kapitels beschrieben und wird im Zusammenhang mit der in dieser Arbeit dargelegten Problemstellung im Kapitel 3.4 nochmals konkretisiert.

Im Hinblick auf die Entwicklung einer Arbeitsdefinition kann also mentale Beanspruchung in Anlehnung an aufmerksamkeitstheoretische Modelle als die Menge der verwendeten kognitiven

Ressourcen beschrieben werden, die für die Bearbeitung einer Aufgabe zu investieren sind (vgl. Norman & Bobrow, 1975; Shiffrin & Schneider, 1977). Dabei ist von verschiedenen Ressourcenreservoirs für die verschiedenen Verarbeitungsstufen des informations-verarbeitenden Systems Mensch auszugehen (Card, Moran, & Newell, 1986; Wickens, 1984). Diese maximale Menge der in den Reservoirs vorhandenen Ressourcen ist allerdings begrenzt. Bei der gleichzeitigen Bearbeitung mehrerer Teilaufgaben können allerdings Interferenzen entstehen, da gleiche Reservoirs zur Lösung mehrerer Aufgaben benötigt werden, so dass mit zunehmender Aufgabenschwierigkeit oder Komplexität die maximal vorhandene Menge an Ressourcen erreicht wird (vgl. Manzey, 1988). Damit ist keine weitere Leistungssteigerung möglich bzw. sogar Leistungsabnahme oder das Auftreten von Fehlern wahrscheinlich, wenn keine Anpassung der Handlungsstrategie vorgenommen wird (Desmond & Hoyes, 1996; Hollnagel, 2009; Loft, Sanderson, Neal, & Mooij, 2007; Sperandio, 1978).

Mentale Beanspruchung kann hierbei definiert werden als: *Die Menge an Informationsverarbeitungskapazität, die für die Aufgabenbearbeitung benötigt werden* (De Waard, 1996) bzw. *der Unterschied zwischen der gesamten verfügbaren Kapazität des Informationsverarbeitenden Systems und der zur Aufgabenausführung benötigten Kapazität* (Gopher & Donchin, 1986). Diese Kapazitätsgrenzen liegen dabei hauptsächlich im Arbeitsgedächtnis begründet. Bei der Bewertung mentaler Beanspruchung wird außerdem davon ausgegangen, dass mit steigenden Anforderungen (und damit sinkenden zur Verfügung stehenden Kapazitäten) die mentale Beanspruchung steigt (siehe z.B. Brooker, 2003).

Eine Erfassung mentaler Beanspruchung kann mit verschiedenen Methoden erfolgen, welche als hinreichend zuverlässig erachtet werden.

2.4 Erfassung Mentaler Beanspruchung und Operateurszustände

Da, wie bereits beschrieben, eine direkte Erfassung mentaler Beanspruchung nicht möglich ist, existiert eine Vielzahl von Methoden, die zumindest indirekt Rückschlüsse auf den Grad der mentalen Beanspruchung zulassen sollen. Hierbei können grundlegend drei große Gruppen von Methoden zur Erfassung mentaler Beanspruchung unterschieden werden: *physiologische*, *subjektive* und *leistungsbasierte Maße* (vgl. Brookhuis & DeWaard, 2001; DeWaard, 1996; Miller, 2001; O'Donnell & Eggemeier, 1986). Diese Methoden werden vor allem in experimentellen Umgebungen eingesetzt, um die mentale Beanspruchung von Operateuren zu erfassen. Darüber hinaus werden sie auch verwendet, um in realen Umgebungen den Zustand des Operateurs in Mensch-Maschine-Systemen zu erfassen (zum Beispiel beim Autofahren zur Erfassung von Müdigkeit oder mentaler Beanspruchung vgl. DeWaard, 1996). In der Flugverkehrskontrolle werden vor allem physiologische Indikatoren als Indizes mentaler Beanspruchung verwendet (z.B. Brookings, Wilson, & Swain, 1996; Metalis, 1991; Vogt, Hagemann, & Kastner, 2006),

Darüber hinaus wird, wie bereits beschrieben, durch modellgetriebene Ansätze und mathematische Indizes versucht, Rückschlüsse auf die mentale Beanspruchung aus objektiv messbaren Variablen zu erhalten (z.B. Athènes et al., 2002; Lee et al., 2007; Pawlak et al., 1996; Schmidt, 1976; Thonrnhill, 1995).

Bei der Auswahl von Erhebungsmethoden mentaler Beanspruchung sollte zunächst auf die Einhaltung verschiedener Gütekriterien geachtet werden. Als wichtigste sind hier Sensitivität und Diagnostizität zu nennen. Es können aber auch Interferenzfreiheit, Breite des Anwendungsbereiches, die Größe der Messintervalle sowie der benötigte Aufwand zur Datenerhebung, -verarbeitung und -auswertung für die Auswahl einer Methode eine Rolle spielen (Manzey, 1998; Miller, 2001; Ribback, 2003) [4].

Die drei Gruppen von Maßen besitzen alle Vor- und Nachteile, so dass es nahezu unmöglich ist, ein perfektes Maß zu finden (Miller, 2001). Daher ist es sinnvoll, mehrere Maße aus mindestens zwei der oben genannten Gruppen anzuwenden, um eine möglichst genaue Erhebung mentaler Beanspruchung zu ermöglichen. Die einzelnen Kategorien werden im Folgenden vorgestellt.

2.4.1 PHYSIOLOGISCHE MAßE

Physiologische Maße verwenden die physischen Reaktionen des Körpers, um das Ausmaß mentaler Beanspruchung objektiv zu messen. Sie beruhen auf der Annahme, dass erhöhte mentale Beanspruchung zu erhöhten physischen Reaktionen des Körpers führen (Moray, 1979). Mit physiologischen Maßen lassen sich kontinuierliche Veränderungen mentaler Beanspruchung sehr gut erfassen. Diese Veränderungen können durch die Herz-, Gehirn-, Atmungs- oder Augenaktivität abgebildet werden. Die Herzaktivität wird über die Herzrate, die Herzratenvariabilität, sowie den Blutdruck gemessen. Zur Messung der Hirnaktivität wird entweder das Elektroenzephalogramm (EEG) oder das Elektrookulogramm (EOG) verwendet. Atmungsaktivität misst die Menge an Luft, die eine Person während einer bestimmten Anzahl von Atemzügen einatmet (Miller, 2001). Die Messung der Augenaktivität beinhaltet vor allem horizontale Augenbewegungen, die Liedschlagfrequenz sowie die Schließintervalle aber auch die Messung des Pupillendurchmessers (Pupilometrie; Beatty, 1982).

All diese Maße haben insbesondere den Vorteil, dass sie interferenzfrei sind, das heißt, sie benötigen keine direkte Reaktion der Versuchspersonen (Miller, 2001). Die Schwierigkeit hierbei besteht allerdings darin, dass der Körper nicht nur auf mentale Beanspruchung physiologisch reagiert, sondern auch beispielsweise auf erhöhte physische Beanspruchung (De Waard, 1996). Daher können Ergebnisse verzerrt werden, wenn eine erhöhte mentale

[4] Eine ausführliche Behandlung der Gütekriterien ist nicht im Sinne dieser Arbeit. Daher sei für weitere Informationen an dieser Stelle auf Manzey (1998) verwiesen.

Beanspruchung auch mit erhöhter körperlicher Aktivität einhergeht. Eine Diagnostizität ist damit nicht bzw. nicht immer gewährleistet.

Das am häufigsten verwendete physiologische Maß ist die Herzaktivität (Manzey, 1998; Miller, 2001; Ribback, 2003). Hierbei werden vor allem die tonische Herzrate und die Herzratenvariabilität bestimmt. Bei Untersuchungen im Feld oder bei feldnahen Bedingungen stellt die Herzrate den Beanspruchungsindikator der Wahl dar, da sie leicht zu erfassen ist (Manzey, 1998; Roscoe, 1992, 1993). Sie wird meist in Herzschlägen pro Minute (beats per minute, bpm) oder dem Abstand zwischen zwei Herzschlägen in Millisekunden (inter beat intervall, IBI, auch Herzperiode genannt) ausgewertet. Die Befundlage zur Sensitivität der Herzrate ist als heterogen zu bezeichnen.

Zahlreiche Befunde konnten zeigen, dass die Herzrate (bpm) bei steigender kognitiver Beanspruchung ebenfalls ansteigt. In verschiedenen Labor- und Feldstudien wurde ein Anstieg der Herzrate unter Belastungsbedingungen festgestellt. So konnten Hörhold (1994), sowie Hörhold und Walschburger (1998) bei Konzentrationsaufgaben nach einem initialen Anstieg der Herzrate eine Abnahme die wiederholten Belastungsbedingungen feststellen (z.B. wiederholte oder länger andauernde Ausführung kognitiver Aufgaben), was als Zeichen einer Adaption der Herzrate an die Belastungsbedingungen interpretiert wurde. Speziell beim Abruf aus dem Gedächtnis ohne Sprechaktivität konnte ein Zusammenhang zwischen der Gedächtnisleistung (als Anzahl der Gedächtnisitems) und der Herzrate festgestellt werden (Hanson, Schellekens, Veldman, & Mulder, 1993; Schwer, Furedy, & Heslegrave, 1984). De Waard, Jessurun, Steyvers und Ragatt (1995), sowie Richter, Wagner, Heger und Wiese (1998) stellten eine Kovariation der Herzrate mit der Schwierigkeit der Fahrstrecke beim Autofahren fest. Auch bei Fluglotsen konnte eine Reihe von Untersuchungen eine Erhöhung der Herzrate in Abhängigkeit der Schwierigkeit der Tätigkeit festgestellt werden (Costa, 1993; Henderson, Bakal, & Dunn, 1990; Laurig, Becker-Biskaborn, & Reiche, 1971; Rohmert & Rutenfranz, 1972; Rose & Fogg, 1993; Smith, 1980; Vogt et al., 2006; Vogt & Kastner, 2001). Allerdings konnte bei der Ausführung von Aufgaben in Flugsimulatoren durch Piloten nicht immer ein Zusammenhang zwischen der Herzrate und der Aufgabenschwierigkeit festgestellt werden (Casali & Wierwille, 1983; Wierwille, Rahimi, & Casali, 1985).

In einer groß angelegten Studie zur Untersuchung u.a. der Sensitivität der Herzrate für mentale Beanspruchung stellten Nickel, Eilers, Seehase und Nachreiner (2002) fest, dass die Herzrate, sowie auch die Herzratenvariabilität *„zwar hinreichend sensitiv* [sind]*, um die Belastungsbedingungen Ruhe vs. Arbeit zu differenzieren“* (S.33), es allerdings in keiner ihrer Studien gelang, *„feiner abgestufte Belastungsgrade zu differenzieren“* (S.33). Für eine ausführliche Übersicht verschiedenster Studien zu diesem Thema und der gefundenen Effekte sei hier auf Ribback (2003) verwiesen.

Es kann festgehalten werden, dass zwar eine gewisse Sensitivität der Herzrate auf mentale Beanspruchung besteht, diese aber höchstwahrscheinlich nur in der Lage ist, eine Ruhe- von einer Belastungsbedingung zu differenzieren. Eine feinere Abstufung von Belastungsgraden erscheint nicht möglich (Nickel et al., 2002). Eine Diagnostizität der Herzrate ist nicht gegeben, da ein Anstieg der Herzrate nicht nur auf einen Anstieg der mentalen Beanspruchung zurückzuführen ist, sondern auch auf physische und emotionale Beanspruchung, welche die Effekte mentaler Beanspruchung völlig überlagern können (Manzey, 1998).

Nichtsdestotrotz stellt die Herzrate auf Grund ihrer einfachen Erhebung und geringer Interferenz ein häufig angewandtes Maß mentaler Beanspruchung dar, insbesondere dann, wenn ein Einfluss anderer Beanspruchungsarten weitestgehend ausgeschlossen werden kann. Daher wird die Herzrate als Vergleichsmaß für das hier zu erstellende Modell herangezogen.

2.4.2 Subjektive Maße

Bei subjektiven[5] Maßen wird davon ausgegangen, dass niemand besser in der Lage ist eine Einschätzung der erlebten Beanspruchung abzugeben, als die Person, die es betrifft (De Waard, 1996). Johannsen et al. (1979) begründet die Verwendung auch mit der einfachen Sichtweise, dass, wenn eine Person eine hohe Beanspruchung empfindet, auch tatsächlich eine hohe Beanspruchung vorhanden ist. Obwohl physiologische Maße meist genauer sind, sind subjektive Maße leichter anzuwenden. Muckler und Seven (1992) sehen die Stärke subjektiver Ratings darin, dass sie individuelle Unterschiede in Fähigkeiten, Zuständen und Ansichten mit in Betracht ziehen und sie daher gerade wegen ihrer Subjektivität so wertvoll sind.

Kritik erfahren subjektive Maße vor allem dahingehend, dass bezweifelt wird, dass Versuchspersonen zu einer Introspektion des eigenen Empfindens mentaler Beanspruchung fähig sind (vgl. Annett, 2002), vor allem da es den Versuchspersonen schwer fallen dürfte, physische und mentale Beanspruchung exakt zu unterscheiden (O'Donnell & Eggemeier, 1986). Nach De Waard (1996) kann davon ausgegangen werden, dass die meisten subjektiven Maße hinreichend sensitiv für Veränderungen in der Beanspruchung sind.

Die subjektiven Bewertungsmethoden sind nochmals in *unidimensionale* und *multidimensionale* Messinstrumente zu unterscheiden. Welche hierbei verwendet werden sollten, hängt stark davon ab, welche Informationen benötigt werden. Unidimensionalen Maßen wird eine höhere Sensitivität hinsichtlich der Belastung zugesprochen. Zudem besitzen unidimensionale Skalen eine relativ geringe Interferenz mit der Hauptaufgabe und sind leicht verständlich für die Versuchsperson. Außerdem sind sie für den Versuchsleiter leicht

[5] Meist werden die hier gemeinten Maße als Selbstreport-Maße (self-report measures) bezeichnet, da prinzipiell auch physiologische Messungen subjektiv sind (De Waard, 1996).

anzuwenden und auszuwerten. Allerdings ist die Diagnostizität bei multidimensionalen Instrumenten als höher zu bewerten (De Waard, 1996).

Eine *unidimensionale Skala*, die sogenannte RSME (Rating scale mental effort) oder SEA-Skala (subjektiv erlebte Anstrengung) wurde von Zijlstra und Van Doorn (1985) entwickelt und von Eilers, Nachreiner und Hänecke (1986) ins Deutsche übersetzt und validiert (siehe Anhang). Hierbei wird die mentale Beanspruchung über ein einzelnes Item erfasst. Die Arbeitsgruppe für Ingenieurpsychologie der Humboldt-Universität zu Berlin modifizierte diese Skala hinsichtlich einiger Gestaltungsmerkmale, wie zum Beispiel dem Hinzufügen eines verbalen Endankers oder der Veränderung der numerischen Anker dahingehend, dass sie als Prozentangaben interpretiert werden können (Kosicki, 2011). Die modifizierte Skala ist außerdem länger und enthält zehn statt sieben verbale Anker. Es wurde allgemein erreicht, dass die Daten der modifizierten Skala gegenüber der originalen eine kleinere (Fehler-) Varianz aufweisen. Die Versuchspersonen stimmen stärker in ihren Urteilen überein, was im Sinne der Erhöhung der Messgenauigkeit ist (Prof. Dr. Hartmut Wandke, persönliche Kommunikation am 26.06.2013). Durch das Hinzufügen eines oberen Ankers und durch die farbliche Markierung wurde anscheinend der subjektive Interpretationsspielraum eingeschränkt (Prof. Dr. Hartmut Wandke, persönliche Kommunikation am 26.06.2013; Kosicki, 2011). Ein Vergleich dieser RSMEmod-Skala mit anderen etablierten Messinstrumenten, wie dem NASA-TLX (Hart & Staveland, 1986) und der originalen SEA-Skala ist in Kosicki (2011) zu finden.

Als *multidimensionales Maß* wird vor allem der NASA-TLX (NASA Task Load Index; Hart & Staveland, 1986) sehr häufig verwendet. Hierbei werden Bewertungen auf verschiedenen Sub-Skalen vorgenommen, womit z.B. geistige Anstrengung und physische Anstrengung unterschiedlich erfasst werden können. Anschließend können diese Bewertungen zusammengefasst werden, um ein Gesamturteil zu erhalten. Die Bestimmung dieses Gesamturteils erfordert beim NASA-TLX allerdings den Vergleich aller sechs Sub-Skalen untereinander, so dass 15 Vergleiche nötig sind, bevor das Gesamturteil bestimmt werden kann, wodurch sich die Auswertung hier etwas umfangreicher gestaltet. Daher schlagen Byers, Bittner und Hill (1989) die Berechnung eines sogenannten Raw Task Load Index (RTLX) vor, der diese Vergleiche nicht erfordert, sondern sich durch Bildung eines einfachen Mittelwertes aller sechs Skalen ergibt. Zusätzlich zur umfangreicheren Auswertung multidimensionaler Verfahren ist hier auch eine stärkere Interferenz mit der Primäraufgabe gegeben, da das Ausfüllen längere Zeit in Anspruch nimmt und damit selbst höhere Aufmerksamkeitsressourcen erfordert.

Auf Grund der einfachen Anwendung und Auswertung unidimensionaler Skalen, sowie deren hoher Sensitivität und geringer Interferenz mit der Hauptaufgabe, ist die RSME-Skala zur Erfassung mentaler Beanspruchung sehr geeignet. Allein die Diagnostizität kann als

Schwachstelle angesehen werden, so dass hier eine sinnvolle Kombination von uni- und multidimensionalen Verfahren anzustreben ist.

2.4.3 Leistungsbasierte Maße

Wie in Kapitel 2.3 beschrieben, halten Hollnagel und Woods (2005) die Erfassung mentaler Beanspruchung nur aufgrund von Rückschlüssen direkter Leistungserfassungen für möglich. Hierbei können Leistungsmaße in der *Primäraufgabe* oder in einer *Sekundäraufgabe* erfasst werden. Die primäre Leistungsmessung erfolgt durch die direkte Messung beispielsweise der Anzahl der gemachten Fehler oder Ausführungs- oder Reaktionszeiten in der Primäraufgabe. Alle Leistungsmaße der Primäraufgabe sind dabei Geschwindigkeits- oder Genauigkeitsmaße (De Waard, 1996). Die Bewertung der Sensitivität solcher Maße ist etwas schwierig, da eine konstante Leistung, z.B. ein Ausbleiben von Fehlern, nicht unbedingt bedeutet, dass ein Operateur auch gering belastet ist. Ebenso können Unterschiede zwischen zwei Operateuren nicht aufgedeckt werden. So kann der eine bereits an der Grenze seiner Leistungsfähigkeit angekommen sein, während ein anderer noch sehr gut in der Lage ist, die Aufgabe zu erfüllen, allerdings dennoch keine Fehler auftreten. Daher schlägt De Waard (1996) die Kombination mehrerer Erfassungsmethoden mentaler Beanspruchung und der Leistungserfassung in der Primäraufgabe vor, um valide Aussagen über die Beanspruchung während der Aufgabenausführung zu machen.

Auf Grund der beschriebenen Probleme bei der Erfassung der Leistung in der Primäraufgabe, ist diese im praktischen Anwendungsbereich eher wenig sinnvoll. In der Flugverkehrskontrolle schließen sich diese Maße mit dem Ziel der gründlichen und gleichzeitig effizienten Abwicklung aus. Verschlechtert sich die Leistung in der Primäraufgabe oder treten sogar Fehler auf, so kann es bereits zu spät sein, um Schlimmeres zu verhindern. Daher können diese Maße lediglich in Simulationsexperimenten Anwendung finden, wenn hierbei alle weiteren Konsequenzen von bspw. Fehlern ausgeschlossen werden können. Als Vergleichsmaß für die Validierung von Modellen in Simulationsexperimenten können solche Maße daher durchaus herangezogen werden.

Eine weitere Möglichkeit der Leistungserfassung ist das Sekundäraufgabenparadigma. Hierbei wird eine Zweitaufgabe zur Primäraufgabe hinzugefügt, um die Erfassung der „freien Kapazität“ zu ermöglichen. Nach dem sogenannten „Subsidiary Task Paradigm“ wird der Zweitaufgabe hierbei eine geringere Priorität zugeordnet als der Primäraufgabe. Die Leistung in dieser Zweitaufgabe verändert sich damit mit steigender Schwierigkeit der Erstaufgabe und weist somit auf die „freie Kapazität“ hin (De Waard, 1996). Hierbei sollte nach der Theorie multipler Ressourcen von Wickens (1984) die Zweitaufgabe möglichst die gleichen Ressourcen beanspruchen, wie die Erstaufgabe und somit eine hohe Interferenz erzeugen. Häufig

verwendete Zweitaufgaben sind beispielsweise Gedächtnisabruf oder das Lösen von Rechenaufgaben. Zur Sensitivität dieser Maße ergibt sich ein eher diffuses Bild (Eggemeier & Wilson, 1991).

Beim Zweitaufgabenparadigma besteht vor allem die Problematik der Interferenz mit der Hauptaufgabe, da durch das Hinzufügen einer weiteren Aufgabe die mentale Beanspruchung künstlich nach oben getrieben wird. Ebenso wie die Messung der Leistung in der Primäraufgabe ist dies für die Praxis wenig sinnvoll. Es kann zwar von einer guten Sensitivität ausgegangen werden, wenn Primär- und Sekundäraufgabe die gleichen Ressourcen beanspruchen. Als problematisch ist allerdings anzusehen, dass die Zweitaufgabe bei hoher Beanspruchung ignoriert werden kann, sowie dass die Ressourcenverteilungsstrategie der Versuchspersonen nicht bekannt ist (De Waard, 1996). Daher sollte, wenn möglich, die Messung der Leistung in der Primäraufgabe vorgezogen werden, da diese sehr viel leichter erfassbar ist (Miller, 2001).

2.4.4 Zusammenfassende Bewertung der Erfassungsmethoden

Mit dem Ziel Warn- oder adaptive Unterstützungssysteme zu entwickeln werden die vorgestellten Methoden vor allem im Bereich der Fahrzeugführung eingesetzt, um Fahrerzustände online erkennen zu können (vgl. Karrer-Gauß, 2012). Hierbei sollten insbesondere die genannten Gütekriterien berücksichtigt werden, um zu prüfen, wie gut eine Erfassungsmethode für die jeweilige Problemstellung von Nutzen sein kann. Beim Autofahren können vor allem physiologische Methoden in Frage kommen, da sie eine relative geringe Interferenz mit der Primäraufgabe aufweisen. Subjektive Maße, sowie das Zweitaufgabenparadigma sind hierbei eher ungeeignet, da sie ein hohes Ablenkungspotential mit sich bringt.

Auch für die Erfassung des Zustandes, vor allem der mentalen Beanspruchung des Lotsen bei seiner Kontrollaufgabe, können die vorgestellten Maße Anwendung finden. Auch hier ist speziell die Interferenzfreiheit zu beachten, da die Flugverkehrskontrolle einen Hochsicherheitsbereich darstellt und somit jegliche Ablenkung zu Fehlern führen könnte. Weiterhin ist im Hinblick auf die Entwicklung adaptiver Unterstützungssysteme für den Lotsen, eine hohe Sensitivität und Diagnostizität zu gewährleisten, da der Automatisierungsgrad und der Grad der Hilfestellung, die das Unterstützungssystem bietet, hier von der aktuellen mentalen Beanspruchung abhängt. In Tabelle 1 sind die vorgestellten Erfassungsmethoden nochmals zusammengefasst. Es wird weiterhin ein Vergleich dieser Methoden hinsichtlich wichtiger Gütekriterien vorgenommen, die bei der Auswahl von Erfassungsmethoden eine Rolle spielen können. Hierbei sind insbesondere subjektive und physiologische Maße zwar als relativ sensitiv zu bewerten, weisen allerdings eine eher geringe Diagnostizität auf. Eine Erfassungsmethode der Wahl sollte vorzugsweise eine hohe Diagnostizität für mentale Beanspruchung gewährleisten, da in der Flugverkehrskontrolle

auch andere Beanspruchungsformen, wie körperliche Beanspruchung, auftreten können. Vor allem das Sprechen kann bei physiologischen Maßen große Probleme bereiten. Auch wenn subjektive Maße hinreichend interferenzfrei und leicht anzuwenden wären, so ist ihre zeitliche Auflösung eher schlecht, so dass unidimensionale Skalen oft hintereinander angewendet werden müssen, um eine höhere zeitliche Auflösung zu erreichen, wodurch eine höhere Interferenz mit der Hauptaufgabe entstehen kann. Diese Interferenz besitzen Sekundäraufgaben-Maße bereits durch die Art der angewendeten Messung, auch wenn sie als hoch sensitiv und diagnostisch für mentale Beanspruchung angesehen werden können.

Tabelle 1: Vergleich der vorgestellten Methoden zur Erfassung mentaler Beanspruchung und Bewertung der Gütekriterien (++ sehr gut; -- sehr schlecht)

Kriterium	**Subjektive Maße**	**Physiologische Maße**	**Primär-aufgabe**	**Sekundär-aufgabe**
Sensitivität	+	+	+	++
Diagnostizität	-	-	+	++
Interferenzfreiheit	+	++ (EKG) + (Pupillometrie) + (spontan EEG)	++	--
Anwendungsbreite	++	++ (EKG) -- (Pupillometrie) + (spontan EEG)	-	-
Leichte Anwendung	++	- (EKG) -- (Pupillometrie) -- (spontan EEG)	+	-
Leichte Auswertung	++	- (EKG) -- (Pupillometrie) -- (spontan EEG)	+	-
Zeitliche Auflösung	-	++ (EKG) ++(Pupillometrie) ++ (spontan EEG)	-	-

Leistungsmessungen in der Primäraufgabe, wie z.B. die Fehlerzahl, sind im Feld als eher ungeeignet zu bezeichnen, da dieses Maß den Zielen der Flugverkehrskontrolle wiederspricht. Zusätzlich sind mit allen Messmethoden keine Prognosen der mentalen Beanspruchung eines Operateurs möglich. Eine solche Einschätzung könnte aber in der Flugverkehrskontrolle beispielsweise eine bessere Verkehrsplanung ermöglichen, indem der Verlauf der mentalen Beanspruchung des Lotsen über die Zeit mit berücksichtigt wird.

2.5 KAPITELZUSAMMENFASSUNG

Als mental beanspruchend werden solche Aufgaben beschrieben, die primär durch geistige Anforderungen ein subjektives Gefühl des Angestrengtseins hervorrufen und damit Anforderungen an verschiedene Teile des menschlichen Informationsverarbeitungssystems stellen. Dabei existiert bisher noch keine einheitliche und umfassende Theorie mentaler Beanspruchung. Vielmehr können zwei dominierende theoretische Ansätze unterschieden werden: aktivierungstheoretische und aufmerksamkeitstheoretische Ansätze (vgl. Kapitel 2.1). Innerhalb der aufmerksamkeitstheoretischen Ansätze wird mentale Beanspruchung als die Menge der verwendeten kognitiven Ressourcen beschrieben, die für die Bearbeitung einer Aufgabe zu investieren sind. Dabei ist von verschiedenen Ressorcenreservoirs für die verschiedenen Verarbeitungsstufen des informationsverarbeitenden Systems Mensch auszugehen. Für die zentrale Verarbeitung von Informationen stellt vor allem die Begrenzung des Arbeitsgedächtnisses *den* limitierenden Faktor dar. Auch hierbei existieren verschiedene Modellvorstellungen über den Aufbau und die Funktionsweise des Arbeitsgedächtnisses, sowie dessen Begrenzung. Unter der Annahme einer Single-Limited Capacity Theorie kann davon ausgegangen werden, dass Lotsen in der Lage sind zehn flugverkehrskontrollspezifische Informationseinheiten (sogenannte Chunks) gleichzeitig im Arbeitsgedächtnis zu speichern (Kapitel 2.2). Da mentale Beanspruchung keiner direkten Messung zugänglich ist, erfolgt in der Anwendungsdomäne Flugverkehrskontrolle der Versuch einer Quantifizierung mentaler Beanspruchung bisher vornehmlich durch die Berechnung mathematischer Indizes (z.B. Traffic Load Index, TLI) auf Grundlage von Belastungsfaktoren. Dabei finden hauptsächlich ressourcentheoretische Modelle mentaler Beanspruchung, sowie das Belastungs-Beanspruchungskonzept nach Rohmert (1984) Anwendung. Kognitive und personenbezogene Faktoren werden dabei eher vernachlässigt, sodass diese Methoden im Hinblick auf das Design adaptiver Automation ein unzureichendes Bild erzeugen (vgl. Kapitel 2.3). Auch andere weithin akzeptierte Methoden zur Erfassung mentaler Beanspruchung, wie physiologische (vgl. Kapitel 2.4.1), subjektive (vgl. Kapitel 2.4.2) und leistungsbezogene Maße (vgl. Kapitel 2.4.3) können verwendet werden. Diese besitzen allerdings ihrerseits einige Defizite hinsichtlich Gütekriterien wie Diagnostizität, Sensitivität und Interferenzfreiheit (Kapitel 2.4.4). Es ist zu erkennen, dass keines der beschriebenen Maße als ideal anzusehen ist. Daher ist, wie De Waard (1996) vorschlägt, eine Kombination verschiedener Maße sehr sinnvoll.

3 STAND DER FORSCHUNG

Im folgenden Kapitel sollen zunächst das Ziel, sowie die Stärken und Schwächen der kognitiven Modellierung in Mensch-Maschine-Systemen erläutert werden. Hierzu erfolgt zunächst allgemein eine Einführung des Begriffes der Modellierung und Simulation in Mensch-Maschine-Systemen. Aus den Ergebnissen des Kapitels 3.1 ergibt sich die Notwendigkeit eines alternativen Ansatzes zur klassischen kognitiven Modellierung. Daher werden in Kapitel 3.2 zwei Forschungsansätze im Bereich Mensch-Maschine-Systeme – der klassische mikrokognitive und der alternative makrokognitive – vorgestellt. Diese Unterscheidung wird auf die kognitive Modellierung und Simulation übertragen. Kapitel 3.3 geht dann auf ausgewählte bereits vorhandene Modellierungsansätze im Bereich Flugverkehrskontrolle und im Speziellen zu mentaler Beanspruchung ein. In Kapitel 3.4 wird aus dem vorgestellten Forschungsstand zu Theorien mentaler Beanspruchung, sowie zur Modellierung in der Flugverkehrskontrolle die Notwendigkeit der Entwicklung eines Modells mentaler Beanspruchung abgeleitet sowie diskutiert, welchen Nutzen ein solches Modell erbringen kann. Kapitel 3.5 fasst das Kapitel zusammen.

3.1 MODELLIERUNG UND SIMULATION VON MENSCH-MASCHINE-SYSTEMEN

Ein System kann im einfachsten Fall nach Dörner (2000) als *„…ein Geflecht von miteinander verknüpften Variablen.“* (S.109) beschrieben werden. Als Mensch-Maschine-Systeme bezeichnet man dabei spezielle Systeme, bei denen das Zusammenwirken mindestens eines Menschen mit einem technischen System zur Aufrechterhaltung der Funktion notwendig ist (z.B. Kraftfahrzeuge, chemische Anlagen, Flugzeuge) (Dzaack, 2008). Nach Timpe und Kolrep (2002) bezeichnet der Begriff Mensch-Maschine-System *„eine zweckmäßige Abstraktion des zielgerichteten Zusammenwirkens von Personen mit technischen Systemen zur Erfüllung eines fremd- oder selbstgestellten Auftrags“* (S. 10).

Die Entwicklung und Analyse neuer Technologien und Systeme erfolgt in den Ingenieurwissenschaften seit jeher mit Hilfe von Modellen und Simulationen und ist hier gängige Praxis. Die Methodenvielfalt reicht dabei von relativ einfachen mathematischen Modellen und Berechnungen, z.B. zur Auslegung von Kühlkreisläufen für einen chemischen Reaktor, die der Ingenieur größtenteils noch mit Bleistift und Taschenrechner durchführen kann, bis zu aufwändigen computerbasierten Verkehrssimulationen, wie zum Beispiel für das Bahnhofsprojekt Stuttgart 21, welche eine Vielzahl von Parametern berücksichtigen, die der Mensch allein nicht mehr überblicken könnte. Dabei können Vorhersagen über das Verhalten eines neuartigen Systems getroffen werden, bevor das System real existiert. Dies ermöglicht vor allem bei sicherheitskritischen Systemen eine einfachere und ungefährlichere Untersuchung

(Möhlenbrink, 2011). Es ist dabei von Vorteil, dass technische Prozesse durch physikalische Gesetzmäßigkeiten mit hinreichender Genauigkeit beschrieben werden können.

Bei Mensch-Maschine-Systemen liegt der spezielle Fall vor, dass zusätzlich zu technischen Komponenten als weitere funktionale Komponente der Mensch Bestandteil des Systems ist und zur Erreichung des fremd- oder selbstgestellten Ziels beiträgt. Es ist eine Kooperation von Mensch und Maschine nötig, um die erwünschte Gesamtleistung zu erreichen. Für die Erreichung einer möglichst ganzheitlichen Modellbildung und/oder Simulation eines solchen Systems muss auch die Komponente Mensch mit in den Modellbildungs- und Simulationsprozess einbezogen werden. Wie Möhlenbrink (2011) beschreibt, wäre dies kein Problem, wenn die Modellierung des Menschen bereits geklärt wäre und nur die Einbindung des Modells des technischen Prozesses zu realisieren wäre. Die Sachlage ist allerdings umgekehrt: Das technische System kann hinreichend auf mathematische Modelle reduziert werden, die Schwierigkeit besteht darin, den Menschen auf modellierbare Funktionen zu reduzieren (Möhlenbrink, 2011). Eine solche Aufgabe ist allerdings nicht trivial, da das Verhalten des Menschen als komplex und inhomogen zu bezeichnen ist (Ruckdeschel, 1997). Daraus ergibt sich unweigerlich die eher philosophische Frage, ob eine solche Reduzierung des Menschen überhaupt angemessen ist.

Die Modellbildung, beschreibt sie nun technische Prozesse oder menschliches Verhalten und kognitive Prozesse, beinhaltet immer die Notwendigkeit einer Komplexitätsreduktion und das Erstellen eines „Abbildes der Realität", nicht aber die Realität selbst. Hierbei werden abstrakte Repräsentationen eines Teils der Realität angestrebt, *„bei denen unwesentliches in Bezug auf den Modellzweck weggelassen wird"* (Leuchter, 2009, S. 6). Ruckdeschel (1997) erläutert dies durch die Einführung mehrerer Begrifflichkeiten. Das reale System ist hierbei der zu modellierende *reale Prozess* wie z.B. ein chemischer Prozess oder der Flughafenprozess. Dieser könnte durch ein sogenanntes *Basismodell* abgebildet werden, welches das vollständige Eingangs-Ausgangs-Verhalten des realen Systems abbildet. Ein solches Basismodell ist als sehr komplex und umfangreich anzusehen, da es aus vielen stark interagierenden Komponenten besteht (Ruckdeschel, 1997). Diese Komplexität verbietet die Verwendung des Basismodells als Simulationsmodell. Nach Festlegung eines *experimentellen Rahmens*, kann allerdings ein wesentlich einfacheres, jedoch innerhalb des experimentellen Rahmens dennoch gültiges, *vereinfachtes Modell* formuliert werden (Ruckdeschel, 1997).

Je nachdem, welcher Zweck die Simulation menschlichen Verhaltens in Mensch-Maschine-Systemen verfolgt, kann eine solche Vereinfachung zunächst dahingehend vorgenommen werden, welcher Teilaspekt menschlichen Verhaltens bei der Interaktion mit einem technischen System abgebildet werden soll. Leuchter (2009) beschreibt hierbei sechs Klassen von Modellen für Nutzereigenschaften bzw. Mensch-Maschine-Interaktion: *Anthropometrische Modelle,* die zur

Untersuchung der physiologischen Beschaffenheit von Bedienern verwendet werden, wobei vor allem ergonomische Aspekte bei der Bedienung (z.B. Kräfte oder Beschleunigungen, denen der Mensch ausgeliefert ist) simuliert werden (Schmidtke, 1976). Sogenannte *Regler-Mensch-Modelle* beruhen auf mathematischer Repräsentation von Regelungsverhalten in hochdynamischen Mensch-Maschine-Systemen, wobei Differentialgleichungen verwendet werden, um die zeitliche Veränderlichkeit dynamischen Verhaltens mathematisch zu formulieren. Es wird versucht das Bedienerverhalten über die Formulierung von Gleichungen so gut wie möglich zu approximieren (vgl. Jürgensohn, 2002). *Aufgabenmodelle* nehmen die Formalisierung einer Arbeitsaufgabe und ihrer Abarbeitung in Form einer hierarchischen Spezifikation erforderlicher Aktivitäten und deren Bedingungen auf kognitiver, motorischer und Wahrnehmungsebene vor. *Konzeptuelle Modelle* beschreiben menschliche Faktoren in Mensch-Maschine-Systemen mit Hilfe von Aufgabenmodellen einerseits auf der Basis qualitativer Modelle menschlichen Verhaltens (z.B. Hacker, 1986; Rasmussen, 1983; Reason, 1991) und anhand von Fehlerwahrscheinlichkeiten für bestimmte Handlungen (Human Reliability Analysis z.B. THERP; Swain & Guttman, 1983; SLIM-MAUD; Embrey, Humphreys, Rosa, Kirwan, & Rea, 1984; CREAM; Hollnagel, 1998). Die Abbildung mentaler Strukturen und Vorgänge ist das Ziel *kognitiver Modelle*. Basierend auf Methoden der symbolischen KI werden kognitionswissenschaftlich fundierte Programme formuliert, mit denen mentale Modelle von Benutzern simuliert werden können (Leuchter, 2009). Repräsentationen des Benutzerwissens, sogenannte *Wissensmodelle*, werden in der Informatik verwendet, um nutzeradaptive Benutzungsschnittstellen zu implementieren. Dabei Erfolgt eine Speicherung des Benutzerwissens über die Aufgabendomäne, Konzepte des technischen Systems und der Benutzungsschnittstelle und der aktuell verfolgten Ziele in Faktenform, um die Darstellung der Dialoge der Software daran anzupassen.

Aus dieser Unterscheidung kann bereits geschlossen werden, dass sowohl anthropometrische Modelle, als auch regelungstechnische Ansätze, für die Zielstellung dieser Arbeit nicht in Frage kommen, da sie nicht in der Lage sind, Wissensstrukturen abzubilden. Aufgaben- und Konzeptuelle Modelle können ihrerseits lediglich als Grundlage für die Implementierung lauffähiger Modelle in Computerprogrammen dienen, da sie eher qualitativen Charakter besitzen. Bei der Simulation von Verhalten in verteilten Mensch-Maschine-Systemen mit hoher Eigendynamik des technischen Prozesses, sowie Problemstellungen aus dem Human Factors Bereich, wie der Simulation mentaler Beanspruchung, ist die Modellierung von Wissensstrukturen essentiell (Leuchter, 2009). Hierzu sind lediglich kognitive und Wissensmodelle in der Lage.

3.2 MIKRO- UND MAKROKOGNITIVE FORSCHUNGS- UND MODELLIERUNGSANSÄTZE

Im speziellen Falle der Modellierung von Mensch-Maschine-Systemen und Human Factors Fragestellungen stellt sich die weitere Herausforderung einer möglichst ganzheitlichen Beschreibung des Mensch-Maschine-Systems, um zu verhindern, dass falsche Einschätzungen über das System getroffen werden (Kindsmüller, Leuchter, Schulze-Kissing, & Urbas, 2004). Es bestehen verschiedene Ansichten darüber, inwieweit ein Basismodell abstrahiert werden sollte bzw. wie feingranular die Beschreibung der Prozesse erfolgen muss, um das vereinfachte Modell zu erhalten. Cacciabue und Hollnagel (1995) unterscheiden hierbei zwei grundlegende Forschungsansätze, welche auch auf die Modellierung übertragen werden können. Diese bezeichnen sie als *mikro-* und *makrokognitiv,* womit sie Ansätze der grundlagenwissenschaftlich orientierten experimentellen Psychologie (mikrokognitiv) von denen der anwendungsorientierten angewandten Ergonomie bzw. den Ingenieurwissenschaften (makrokognitiv) unterscheiden.

3.2.1 MIKROKOGNITIVE ANSÄTZE UND KOGNITIVE ARCHITEKTUREN

Als mikrokognitiv beschreiben Cacciabue und Hollnagel (1995) vornehmlich Erklärungsansätze, die aus der experimentellen Psychologie stammen und eine möglichst detaillierte, theoretische Beschreibung des Funktionierens kognitiver Prozesse bei der Aufgabenbearbeitung zum Ziel haben. Dabei werden Theorien zu spezifischen Phänomenen erstellt und diese mit empirischen Daten korreliert. Nachteile sehen Cacciabue und Hollnagel (1995) vor allem darin, dass der Schwerpunkt auf experimenteller Kontrolle, nicht aber auf externer Validität liegt (vgl. Klein et al., 2003). Das bedeutet, dass Phänomene, die im Labor beobachtbar auftreten und Verhalten signifikant beeinflussen, in realen Umgebungen vernachlässigbar klein sein können. Sie argumentieren, dass Mikrokognition die Verbindung zwischen dem Phänomen und dem realen Kontext zu Gunsten der Übereinstimmung mit der Theorie oder dem Modell aufgeben. Hochkomplexe Verhaltensweisen sind oft nur schwer bis überhaupt nicht experimentell zu untersuchen, so dass sehr feingranulare Untersuchungen oft dazu führen, dass immer mehr Wissen über immer kleinere Problemstellungen generiert wird *(„dass wir immer mehr über immer weniger wissen";* Kindsmüller et al., 2004, S. 10).

Kognitive Modelle im Bereich der mikrokognitiven Ansätze können nach Salvucci und Lee (2003) nochmals in High- und Low-Level Ansätze unterschieden werden (vgl. Abbildung 3). *High-level Ansätze* beschreiben das Nutzerverhalten durch einfache grundlegende Nutzerhandlungen (z.B. eine Mausbewegung oder einen Tastendruck). Das Verhalten von Nutzern wird durch *Aufgabenmodelle* als lineare Sequenz einzelner Schritte beschrieben. Die Methode *GOMS* (Goals, Operators, Methods and Selection Rules) und deren Derivate sind hier die wichtigsten Vertreter (z.B. Card, Moran, & Newell, 1983; Gray, John, & Atwood, 1992; John &

Kieras, 1994; Kieras, 1996; Urbas, Heinath, Schaft, & Leuchter, 2007; Urbas & Leuchter, 2008; Wandmacher, 2002). *Low-level Ansätze* werden auf Basis sogenannter *kognitiver Architekturen* realisiert, welche aus kognitionswissenschaftlichen Theorien entstanden und Annahmen über die Funktionsweise der grundlegenden Komponenten des menschlichen Geistes im Sinne einer „Unified Theory of Cognition" (Newell, 1990) mit Hilfe von Programmiersprachen als *kognitives Modell* implementieren.

Hierbei wird davon ausgegangen, dass intelligentes Verhalten etwas Berechenbares ist und somit nur die richtigen Datenstrukturen und Algorithmen zu finden sind, um das gewünschte Verhalten erzeugen zu können. Nach Langley, Laird und Rogers (2006) beschreibt eine kognitive Architektur dabei die zugrundeliegende Infrastruktur für ein intelligentes System und beinhaltet diejenigen Aspekte eines kognitiven Agenten, die über die Zeit und über verschiedene Anwendungsdomänen konstant sind. Promiente Vertreter kognitiver Architekturen sind beispielsweise ACT-R (Adaptive Control of Thought – Rational; Anderson & Lebiere, 1998; Anderson et al., 2004), SOAR (State, Operator And Result; Laird, Newell, & Rosenbloom, 1987; Laird, 2012), EPIC (Executive Processes, Interactive Control; Kieras & Meyer, 1997; Meyer & Kieras, 1997) oder CLARION (Connectionist Learning with Adaptive Rule Induction ON-line; Sun, 2006).

Den wohl prominentesten und am weitesten verbreiteten und verwendeten Vertreter stellt hierbei ACT-R dar. Die Architektur zielt dabei darauf ab, die menschliche Kognition und Leistung mit einer sehr genauen zeitlichen Auflösung zu modellieren (Möhlenbrink, 2011). Dabei stellt ACT-R eine hybride kognitive Architektur dar, die symbolverarbeitende Struktur durch ein Produktionensystem (anhand Inferenzregeln) und subsymbolische Struktur durch eine Vielzahl an parallelen Prozessen abbildet, wovon letztere durch mathematische Gleichungen beschreibbar sind (Dzaack, 2008). Die grundlegende Struktur von ACT-R besteht aus einer Reihe von Modulen und vorgelagerten Puffern, die verschiedene Arten von Informationen speichern und verarbeiten (z.B. visuelles Modul für die Ausrichtung visueller Aufmerksamkeit, manuelles Modul zur Kontrolle motorischer Handlungen). Wissen in Form von Fakten ist in ACT-R im deklarativen Speicher abgelegt. Über die sogenannte Aktivierung der Wissenseinheiten, wird zu jeder Zeit bestimmt, ob und wie schnell dieses Wissen abgerufen werden kann. Das abgerufene Wissen steht dann dem Arbeitsspeicher zur Verfügung. Dem deklarativen Wissen steht das prozedurale Wissen gegenüber, welches Handlungswissen in Form von Regeln enthält. Sie bilden mentale Verarbeitungsmechanismen ab und dienen der Zielverfolgung (Möhlenbrink, 2011)

Von einem stark anwendungsorientierten Standpunkt aus betrachtet, besteht der Nachteil der Verwendung kognitiver Architekturen dabei insbesondere darin, dass einerseits häufig der Kontext, in dem Kognition stattfindet, vernachlässigt oder nur auf bestimmte relativ abstrakte

Aufgaben reduziert wird, die häufig wenig repräsentativ für reale Situationen sein können. In ACT-R war es ursprünglich z.B. nicht vorgesehen, dass ein erstelltes Modell mit einer simulierten Aufgabenumgebung interagieren kann. Erst durch die Erweiterung ACT-R/PM (perceptual motor) konnten perzeptuelle und motorische Komponenten Berücksichtigt werden (Jürgensohn, 2002). Dabei bleibt allerdings dennoch die Anbindung an simulierte Aufgabenumgebungen bis heute eine schwierige Aufgabe (z.B. Büttner, 2010; Halbrügge, 2013; Hope, Schoelles, & Gray, 2013). Es gilt außerdem zu beachten, dass in Isolation oder zumindest unter experimenteller Kontrolle beobachtbare Phänomene in realen Situationen durchaus irrelevant sein können, da andere viel vorrangiger sind und gefundene Effekte somit nicht generalisierbar sind. Andererseits sind kognitive Architekturen für viele Anwendungsfälle zu feingranular und detailliert, da hier vor allem die „Sichtbarmachung" der kognitive Prozesse im Fokus steht, so dass der Anwendungszweck häufig in den Hintergrund tritt. Zwar werden mit der Modellierung auf Grundlage kognitiver Architekturen auch anwendungsorientierte Problemstellungen adressiert, müssen allerdings sowohl zur Modellierung als auch zur experimentellen Untersuchung wiederum durch starke Vereinfachung und Abstraktion der Aufgabenumgebung erfolgen.

3.2.2 Makrokognitive Ansätze

Als alternativen Ansatz sehen Cacciabue und Hollnagel (1995) den makrokognitiven Ansatz, der im Zuge des Cognitive Systems Engineering (Hollnagel & Woods, 1983) entstand und eher die anwendungsorientierte bzw. ingenieurwissenschaftliche Herangehensweise widerspiegelt. Anspruch dieses Ansatzes ist es, sich mit kognitiven Prozessen in realistischen Aufgaben und unter realistischen Bedingungen zu beschäftigen. Dabei wird weniger Wert auf Phänomene gelegt, die innerhalb des menschlichen Geistes oder ohne offensichtliche Handlung ablaufen. Vielmehr wird menschliche Leistungsfähigkeit unter tatsächlichen Arbeitsbedingungen in den Fokus gerückt und weniger kontrollierte Laborexperimente. Makrokognition ist aber ebenso wenig mit dem behavioristischen „Black Box"-Ansatz zu vergleichen. Vielmehr wird versucht die „Black Box" zu erklären, allerdings nur so weit, wie sie zu erklären ist und wie sie für die vorliegende Problemstellung erklärt werden muss (Cacciabue, 1998b). Einige Phänomene können allerdings sowohl Makro- als auch Mikrokognition zugeordnet werden, wie beispielsweise Problemlösung, Entscheidungsfindung oder Kommunikation. Cacciabue und Hollnagel (1995) beschreiben den Fokus makrokognitiver Modellierung folgendermaßen: „*the interest lies more on how they* [die Phänomene] *are performed and how well they serve to achieve their goals than on the details of what goes on in the human mind while they are performed*" (S. 57).

Im Bereich Mensch-Maschine-Systeme ist die Modellierung in makrokognitiver Tradition eher wenig verbreitet und wird auch weniger häufig angewendet, was unter anderem daran liegen mag, dass hier kein Pendant zu mikrokognitiven Architekturen im Sinne einer „Unified Theorie of Macrocognition“ existiert, sondern kognitive Modelle immer neu, auf das spezifische Problem zugeschnitten erstellt werden müssen. Dies birgt gleichzeitig Vor- und Nachteile. Während viele mikrokognitive Modelle und Simulationen verschiedener Problemstellungen eine gemeinsame Basis, nämlich eine kognitive Architektur besitzen, die auf der Grundlage kognitionswissenschaftlicher Theorien und empirisch belegter Annahmen über die menschliche Informationsverarbeitung erstellt wurde, so sind makrokognitive Modelle und Simulationen immer problemspezifisch und müssen damit immer wieder neu aus der spezifischen Problemstellung heraus erstellt und für das spezifische praktische Problem entwickelt werden. Die Herausforderung besteht daher darin, eine genügend große Korrespondenz zwischen den bekannten oder beobachteten Regularien des Zielphänomens und den Ergebnissen der Simulation zu erhalten oder besser gesagt *„the simulation must be isomorphic to the phenomenon being modeled“* (Cacciabue & Hollnagel, 1995, S. 58). Auch existiert keine gemeinsame Forscher-Community, wie es beispielsweise bei ACT-R der Fall ist. Nicht zuletzt ist dies aber auch der großen Methodenvielfalt geschuldet und dem Fehlen eines einheitlichen theoretischen Fundaments. Als Vorteil kann hierbei allerdings die Tatsache angesehen werden, dass die Modellierung nach makrokognitiven Gesichtspunkten dem Modellierer prinzipiell die Möglichkeit eröffnet, aus dem gesamten Methodenspektrum der Ingenieurwissenschaften und Informatik, aber auch aus der Psychologie und der Kognitionswissenschaft ein für die spezifische Problemstellung geeignetes Beschreibungsmittel auszuwählen.

Der Anspruch, den Cognitive Systems Eingineering dabei erhebt, ist ebenso ein ganzheitlicher. Während mikrokognitive Ansätze häufig auf die Erklärungen von Einzelphänomenen fokussieren und diese experimentell unter stark kontrollierten Bedingungen untersuchen, hat Makrokognition den Anspruch einer ganzheitlichen Systembeschreibung und einer ganzheitlichen Modellierung der Systemperformanz. Hierbei werden kognitive Prozesse nur da betrachtet wo sie nötig sind, wobei eher ein funktioneller, weniger ein struktureller Ansatz verfolgt wird. Ziel des makrokognitiven Ansatzes und damit des Cognitive Systems Engineering ist also die Betrachtung realistischer Phänomene unter realistischen Bedingungen, was bedeutet, dass Untersuchungen möglichst im Feld durchgeführt werden sollten oder wenn nicht anders möglich, im Simulator, jedoch nicht mit Studenten als Probanden, sondern möglichst mit Domainexperten. Ziel der makrokognitiven Modellierung ist dabei nicht die Untersuchung von Kausalzusammenhängen unter Ausschluss möglichst aller Störvariablen, wie es die experimentelle Psychologie tut. Vielmehr wird angestrebt, die menschliche Verarbeitungsfähigkeit bei der Systemgestaltung mit zu berücksichtigen. Nachteilig ist dabei

allerdings, die bereits angesprochene Problemspezifität: Eine getätigte Untersuchung oder Modellierung ist nur unter exakt den gleichen Bedingungen gültig, unter denen sie auch durchgeführt wurde und nur für das spezifische Problem. Sie ist also 1. Nicht auf andere Domänen generalisierbar und muss daher 2. für differierende Bedingungen erneut durchgeführt werden. Um den Ansatz auch begrifflich von der klassischen mikrokognitiven Modellierung abzugrenzen, bezeichnet Cacciabue (1998) makrokognitive Modelle als *kognitive Simulationen*.

Die Abbildung 3 gibt eine Übersicht über die beschriebenen Modellierungsansätze und ausgewählte Beispiele für Modellierungsmethoden bzw. Beschreibungsmittel.

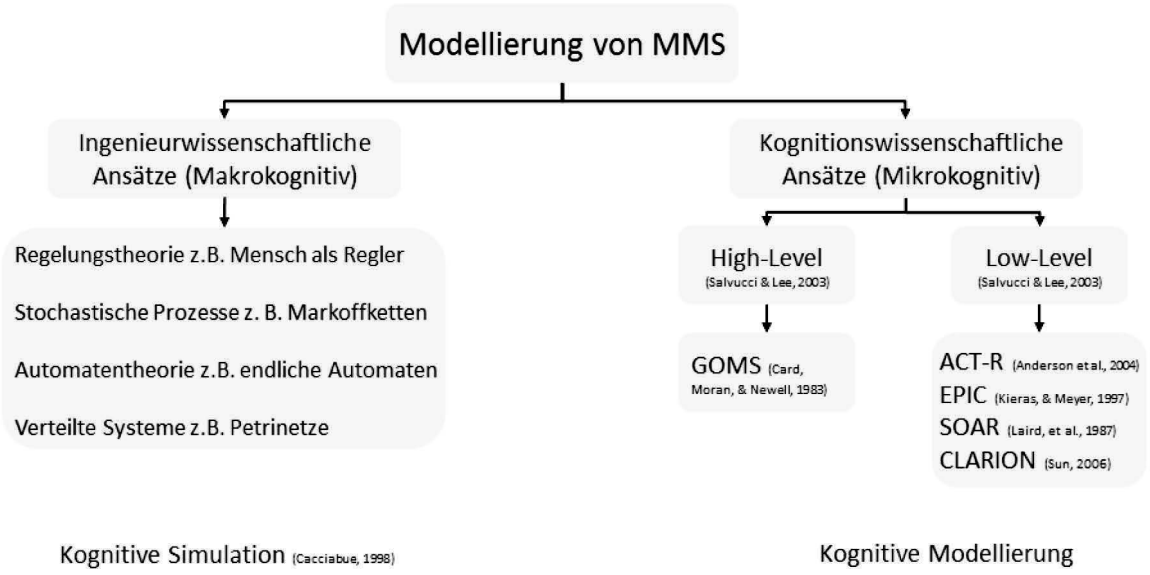

Abbildung 3: Übersicht der Ansätze und zugehöriger Werkzeuge zur Modellierung von Mensch-Maschine-Systemen (eigene Darstellung)

3.2.3 Die Verbindung Mikro- und Makrokognitiver Ansätze

Die durch Cacciabue und Hollnagel (1995) etwas populistisch formulierte und strikte Trennung mikro- und makrokognitiver Ansätze löst allerdings auch einige Skepsis an der Nützlichkeit einer solchen Unterteilung aus. So gibt Flach (2008) zu bedenken: *„with `macrocognition´ we have invented yet another name for a bird that we still know precious little about."* (S.27). Er impliziert, dass die Unterscheidung einfach nur aus verschiedenen Sichtweisen auf ein und dasselbe Phänomen erwächst: *„Do the different constructs represent distinct phenomena, or simply different perspectives on a single phenomenon?"* (S.27). Eine Unterscheidung sollte dabei also keinesfalls so verstanden werden, dass die eine oder die andere Herangehensweise der Forschung im Allgemeinen besser oder schlechter ist, als die andere. Es soll aber dafür sensibilisiert werden, welche unterschiedlichen Sichtweisen sie mit sich bringen und welchem spezifischen Zweck sie damit dienlich sein können. Es ist wichtig, keine Lücken entstehen zu lassen und zu erkennen, dass mikro- als auch makrokognitive Ansätze wichtige Beiträge zum wissenschaftlichen Erkenntnisgewinn liefern und sie daher weniger als konkurrierend oder strikt voneinander getrennt gesehen, sondern zielgenau eingesetzt werden

sollten und sich so auch durchaus sehr gut ergänzen können: *„Rather than differentiating the problems of macrocognition as a special form of cognition, we should be looking for the common threads arnong the different phenomena associated with human experience."* (Flach, 2008, S.38).

Bereits Klein, Klein und Klein (2000) beschreiben die Notwendigkeit der Verbindung von Grundlagenforschung der kognitiven Psychologie (mikrokognitiv) und der Anwendungsforschung der kognitiven Ergonomie (makrokognitiv). Sie beschreiben zwei Möglichkeiten diese Verbindung zu erreichen: Einerseits durch hinreichend große Überschneidung der untersuchten Phänomene, andererseits durch die Anwendung mikrokognitiver Forschungsergebnisse auf anwendungsorientierte makrokognitive Problemstellungen (Klein et al., 2000).

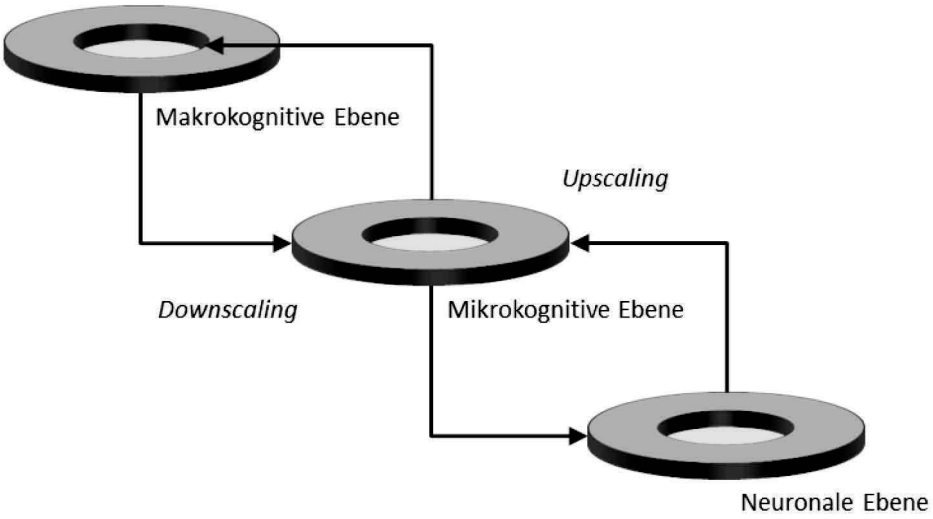

Abbildung 4: Verbindung zwischen makro-, mikrokognitiver und neuronaler Ebene durch Up- und Downscaling (eigene Darstellung nach West et al. 2013).

Für die Modellierung von Mensch-Maschine-Systemen im Speziellen vergleichen Smieszek und Rußwinkel (2013) in Anlehnung an West, Hancock, Somers, MacDougall und Jeanson (2013), den Unterschied zwischen mikro- und makrokognitiver Modellierung mit dem Unterschied zwischen neuronaler und mikrokognitiver Modellierung (vgl. Abbildung 4). Die drei in Abbildung 4 gezeigten Ebenen (makrokognitiv, mikrokognitiv und neuronal) sollten dabei allerdings weniger als strikt getrennt voneinander, sondern eher als ein Kontinuum von stark grundlagenorientierter neuronaler über mikrokognitive hin zu anwendungsorientierter makrokognitiver Modellierung angesehen werden. Somit können durch *Downscaling* mikrokognitive Funktionen in makrokognitive Modelle eingebracht werden, womit eine größere Korrespondenz anwendungsorientierter Modelle mit kognitionspsychologischen Theorien erreicht wird. Umgekehrt können beim *Upscaling* makrokognitive Funktionen und anwendungsorientierte Problemstellungen auf der Grundlage mikrokognitiver Architekturen

modelliert werden (vgl. West et al., 2013), um eine stärkere Anwendungsorientierung zu erreichen. Somit stellen Mikro- und Makrokognition nicht nur verschiedene Ebenen im Kontinuum der Betrachtung menschlicher Kognition dar und dienen unterschiedlichen Zwecken der Mensch-Maschine-Modellierung, sondern können sogar so miteinander verbunden werden, dass Erkenntnisse des einen Ansatzes neue Fragestellungen aufwerfen können, zu deren Lösung der jeweils andere Ansatz beitragen kann (vgl. Smieszek & Joeres, 2013; Smieszek & Rußwinkel, 2013; West & Pronovost, 2009). Im Falle mentaler Beanspruchung ist eine solche Verbindung nicht nur sinnvoll sondern ebenso nötig, da nicht nur äußere Systemfaktoren (makrokognitiv) sondern auch innere kognitive Variablen (mikrokognitiv) das Konstrukt beeinflussen.

So wird mit dieser Arbeit eine Verbindung zwischen mikro- und makrokognitiven Ansätzen angestrebt. Dies stellt in gewisser Weise zwar eine Herausforderung dar, birgt aber die Möglichkeit, reale Phänomene möglichst realistisch z.B. durch ganzheitliche, makrokognitive Modelle abzubilden, um möglichst extern valide Vorhersagen zu erhalten. Gleichzeitig sind die mikrokognitiven Phänomene von Interesse (hier die mentale Beanspruchung) genauer nachzubilden und zu untersuchen um Vorhersagen treffen zu können.

Konkret bedeutet dies, dass ein Modell der Flugverkehrskontrolle an Flughäfen, welches in dieser Arbeit entwickelt werden soll, sowohl den Fluglotsen, als auch dessen zu kontrollierendes System, also den Flughafen, sowie deren Interaktionen nachbilden muss. Dabei müssen nicht nur die inneren „mentalen“ Prozesse des Fluglotsen beschrieben werden, sondern es muss auch bekannt sein, welche Beschränkungen und Regulierungen das Flugverkehrskontrollsystem selbst mit sich bringt. Das bedeutet, sowohl die Aufgaben des Lotsen, die Beschränkungen durch Regularien, die Funktionsweise eines Flughafens, der verwendeten technischen Systeme als auch der kognitiven Prozesse selbst müssen bekannt sein und in den Modellbildungsprozess mit einfließen. Es kann hierbei einerseits auf Methoden zurückgegriffen werden, die als makrokognitiv bezeichnet werden können (z.B. Feldbeobachtungen, Aufgabenanalysen, Analyse von Regularien, Expertenbefragungen und –interviews). Andererseits müssen auch Erkenntnisse, die in experimentellen Settings über die Funktion kognitive Prozesse des Menschen gewonnen wurden mit einbezogen werden (z.B. Informationsverarbeitungskapazität des Menschen, Aufmerksamkeits- und Handlungsmodelle). Zwar ordnet sich die vorliegende Arbeit eher den makrokognitiven Ansätzen zu, schafft aber wie beschrieben auch eine Verbindung beider Ansätze, was so bisher noch nicht erfolgte. Das folgende Kapitel gibt einen Überblick über den Status-Quo im Bereich Modellierung von Mensch-Maschine-Systemen und stellt ausgewählte Modelle mit besonderem Fokus auf die Flugverkehrskontrolle und mentale Beanspruchung vor.

3.3 Ausgewählte Modelle und Simulationen von Problemstellungen im Mensch-Maschine-System Luftfahrt und Flugverkehrskontrolle

Die Forschung und Entwicklung im Bereich der Modellierung und Simulation in Mensch-Maschine-Systemen brachte bereits zahlreiche Modelle hervor, die menschliches Verhalten und kognitive Prozesse sowie verschiedene andere Problemstellungen im Bereich Luftfahrt und speziell den Bereich der Flugverkehrskontrolle adressieren. Von diesen sollen im Folgenden einige ausgewählte vorgestellt werden. Hierbei werden sowohl makrokognitive als auch mikrokognitive Ansätze berücksichtigt. Insbesondere sollen Ansätze beachtet werden, die im speziellen die Problemstellung der Quantifizierung mentaler Beanspruchung zum Ziel haben.

Makrokognitive Modelle bedienen sich vor allem ingenieurwissenschaftlichen Methoden oder Methoden aus der Informatik (z.B. Differentialgleichungen, Petrinetze) zur Modellierung einerseits physikalisch-technischer Problemstellungen aber auch menschlichen Verhaltens und kognitiver Prozesse, wobei vor allem Aufgaben- und konzeptuelle Modelle erstellt werden. Kovács, Németh und Hangos (2005) entwickelten beispielsweise ein Modell auf der Basis farbiger Petrinetze (vgl. Kapitel 4.6) um die Auswirkungen der Verfügbarkeit von Rollwegen auf die Kapazität der Start-Landebahn zu untersuchen, wobei im Modell grundlegende Prinzipien der Flugverkehrskontrolle berücksichtigt werden. Das Modell wurde mit Hilfe des Programmes Design/CPN implementiert (der Vorgänger von CPN-Tools; vgl. Kapitel 4.6) und kann verwendet werden, um optimale Planungssequenzen zu untersuchen. Mit einer ähnlichen Problemstellung beschäftigten sich Vidosavljević und Toŝić (2010). Sie entwickelten ebenfalls ein Modell auf der Basis farbiger Petrinetze, welches zur Untersuchung und Optimierung von Turnaround-Prozessen dient. Hierbei soll der Einfluss verschiedener Strategien zur Ressourcenallokation (Flugzeug-Parkpositionen, Ausrüstung, Personal) auf die Effizienz von Turnaround-Prozessen untersucht werden, um sie so effizient und verlässlich wie möglich gestalten und damit Kosten sparen zu können. Das Modell wurde mit CPN-Tools implementiert.

Diese eher einfachen, auf mehr oder weniger technisch-physikalische Prozesse fixierten Modelle werden durch einerseits konzeptuelle und mathematische Modelle zur Untersuchung menschlicher Zuverlässigkeit (z.B. Daams, Blom, & Nijhuis, 2001) und der Widerstandsfähigkeit (resilience; vgl. Hollnagel, Woods, & Leveson, 2006) komplexer soziotechnischer Systeme (z.B. Stroeve, Everdij, & Blom, 2011) und andererseits durch ganzheitliche Mensch-Maschine-Modelle unter Verwendung farbiger Petrinetze ergänzt. Zunächst zielte die Modellbildung unter Verwendung von Petrinetzen nicht auf die empirische Untersuchung von Verhaltensdaten ab, sondern auf die formale Analyse dieser Modelle für die Arbeitsanalyse und den Designprozess (Möhlenbrink, 2011) oder die Erstellung von Expertensystemen. So beschreibt Ruckdeschel (1997) ein Modell zur Beschreibung regelbasierten Pilotenverhaltens als Teil des Cockpit-Assistenzsystems CASSY (**C**ockpit **Ass**istant **Sy**stem). Werther (2006b) beschreibt erstmalig die

Verwendung farbiger Petrinetze für die kognitive Modellierung und Simulation als Alternative zu kognitiven Architekturen und verfolgt das Ziel, die grundlegenden Komponenten kognitiver Architekturen in einem Petrinetz-Modell mit zu berücksichtigen (vgl. Möhlenbrink, 2011). Das von ihm entwickelte, ganzheitliche Modell (FCR-Modell; Formales Kognitives Ressourcen-Modell) berücksichtigt dabei vor allem die beschränkte Verarbeitungskapazität menschlicher Bediener in Mensch-Maschine-Systemen. Das Modell basiert auf vier Modulationsparametern: Motivstärke, Aktivierungsrate, Kompetenz und Barrierehöhe[6], die bei der Abarbeitung von Handlungszielen die Art und Weise, wie diese abgearbeitet werden, bestimmen. Hierbei liegt die Annahme zugrunde, dass Leistungsunterschiede in der Aufgabenbearbeitung mit den unterschiedlichen Ausprägungen dieser Parameter verbunden sind. Die Validierung des Modellansatzes erfolgt mit Hilfe der Mikrowelt CAMS (Cabin Air Management System; Wastell, Sauer, & Hockey, 2003).

Weiterhin entwickelt Werther (2006a) ein formales Modell der Flughafenverkehrskontrolle auf der Grundlage einer kognitiven Arbeitsanalyse (Vicente, 1999). In Verbindung mit der sogenannten Remote Tower Operation kleinerer Flughäfen (Fürstenau et al., 2009; Oehme & Schulz-Rueckert, 2010; Schmidt, Rudolph, Werther, & Fürstenau, 2006) standen hierbei vor allem die Untersuchung verschiedener Möglichkeiten der Arbeitsorganisation, sowie das Design einer neuen Arbeitsposition (Remote Controller) basierend auf formaler Analyse im Fokus der Modellbildung. Möhlenbrink (2011) verwendet eine Erweiterung dieses Lotsenmodells zur Untersuchung von Entscheidungsheuristiken von Versuchspersonen in einer simulierten Mikrowelt eines Towerlotsenarbeitsplatzes. Ebenso zeigten Möhlenbrink, Manske und Kirlik (2012), wie das 3-Schritt-Prinzip der Heuristischen Entscheidungsfindung (Gigerenzer, Todd, & ABC Research Group, 1999) dazu verwendet werden kann, zu analysieren, wie das Design eines Entscheidungsunterstützungssystems das Entscheidungsverhalten eines Lotsen beeinflusst.

Die vorgestellten ingenieurwissenschaftlich geprägten Modelle beschäftigen sich einerseits mit der Optimierung physikalisch-technischer Prozesse des Flughafens, andererseits mit der Erstellung konzeptueller Modelle menschlicher Zuverlässigkeit und der Widerstandsfähigkeit komplexer soziotechnischer Systeme, sowie mit der Erstellung formaler Modelle der Flugverkehrskontrolle. Hierbei werden kognitive Komponenten oder Prozesse nicht explizit berücksichtigt. Die Arbeitsgruppe um Möhlenbrink beschäftigt sich vor allem mit der Untersuchung heuristischer Entscheidungsfindung mit Hilfe formaler Modelle der Flugverkehrskontrolle. Hierbei ist der Anspruch ein ganzheitlicher, allerdings weniger in Bezug

[6] Motivstärke ist als Abweichung zwischen Ist- und Sollzustand beschrieben. Aktivierungsrate repräsentiert die Energie die das Kognitive System in die Erledigung von Aufträgen investiert und ist mit Arousal vergleichbar. Kompetenz repräsentiert die vorhandenen kognitiven Fertigkeiten. Barrierehöhe repräsentiert das allgemeine Vertrauen in den Erfolg der aktuell auszuführenden Aufgaben (vgl. Werther, 2006b, S. 61f).

auf kognitive Prozesse, als auf die Effizienz der Abarbeitung des Flugverkehres durch den Lotsen unter Berücksichtigung bestimmter Entscheidungshierarchien fokussiert. Lediglich Werther (2006b) versucht die Erstellung einer kognitiven Architektur auf der Basis farbiger Petrinetze, wobei anzumerken ist, dass mit Hilfe einer Anpassung der vier Modulationsparameter Motivstärke, Aktivierungsrate, Kompetenz und Barrierehöhe des FCR-Modells prinzipiell die Erzeugung und Erklärung jeglichen Verhaltens mit Hilfe des Modells möglich ist (overfitting; vgl. Kapitel 4.1). Ferner geben auch Werther und Möhlenbrink zu einem gewissen Grade den Anwendungsbezug zugunsten des Detailgrades der kognitiven Prozesse bzw. der experimentellen Kontrolle (durch Verwendung von Mikroweltsimulationen) auf (mikrokognitiv).

Mikrokognitive Modelle sind im Bereich der Flugverkehrskontrolle vornehmlich für den en-route Bereich erstellt worden. So beschreiben Niessen, Leuchter und Eyferth (1998) ein Modell des mentalen Bildes von Radarlotsen auf der Basis von ACT-R welches mit MoFL (Modell der Fluglotsenleistung) bezeichnet wurde (vgl. auch Jürgensohn, Niessen, & Leuchter, 2002; Niessen, Eyferth, & Bierwagen, 1999; Niessen & Eyferth, 2001). Ziel war die modellhafte Beschreibung und Analyse der kognitiven Prozesse und der resultierenden mentalen Repräsentation einer hochdynamischen Umgebung als Basis für die Konflikterkennung. Pompanon und Raufaste (2009) vergleichen dieses Modell mit einem weiteren Modell von Rantanen und Nunes (2005), welches andere Annahmen bezüglich der Entscheidungen und der Konflikterkennung von Lotsen im en-route Kontext zugrunde legt. Hierbei fiel auf, dass das MoFL-Modell für die Konflikterkennung zunächst die Flugbahnen zweier Flugzeuge vergleicht und festlegt, ob sich diese überschneiden oder nicht. Bei sich überschneidenden Flugbahnen wird die Flughöhe nicht berücksichtigt und sofort die Entscheidung getroffen, dass Konfliktpotential vorliegt. Das Modell von Rantanen und Nunes (2005) hingegen vergleicht zunächst die Flughöhe und prüft nur dann die Trajektorien, wenn gleiche Flughöhen vorliegen. In Experimenten fanden Pompanon und Raufaste (2009) heraus, dass keines der beiden zuvor genannten Modelle alle von ihnen erhobenen Daten hinreichend beschrieb und formulierten ein eigenes Modell, welches hinsichtlich der verwendeten Strategie zur Konflikterkennung als Integration der beiden vorher genannten interpretiert werden kann (Raufaste, 2006). Hierbei wird zunächst die Richtung betrachtet und einer von vier Kategorien zugeordnet. Unmittelbar im Anschluss daran wird überprüft, ob sich die Flughöhe um mehr als 100 Fuß unterscheidet. Beide Informationen werden verwendet, um zu entscheiden, ob ein Konflikt besteht oder nicht.

Ebenfalls auf Basis von ACT-R entwickelte Taatgen (2001, 2002) ein Modell zur Beschreibung individueller Unterschiede beim Erwerb von Fertigkeiten in einer Fluglotsenaufgabe (Kanfer-Ackerman air traffic control task). Hierbei steht allerdings weniger die Vorhersage realer Lotsendaten als vielmehr die Überprüfung einer Theorie individueller Unterschiede beim

Fertigkeitserwerb (Ackerman, 1988, 1990) im Vordergrund. Es wurde davon ausgegangen, dass drei Faktoren – allgemeine Intelligenz, Wahrnehmungsgeschwindigkeit sowie psychomotorische Fähigkeiten – unterschiedliche Teile des Fertigkeitserwerbs beeinflussen. Das Modell verwendete Taatgen (2001, 2002) zur Replikation von Daten, die Ackerman (1988, 1990) zur Aufstellung seiner Theorie erhob und konnte diese damit untermauern. Beim Kanfer-Ackerman air traffic control task handelt es sich um eine stark vereinfachte Fluglotsenaufgabe, die das fehlerfreie Landen von Flugzeugen zum Ziel hat und damit vergleichbar mit der Aufgabe von Towerlotsen ist (vgl. Kapitel 4.3 und 4.4). Sie ist allerdings recht abstrakt repräsentiert und nur über Tastatureingaben zu steuern[7].

Ebenso, wie die Untersuchung menschlicher Entscheidungsfindung, des Fertigkeitserwerbs und der Konflikterkennung, ist auch das Konstrukt mentaler Beanspruchung in der Modellbildung bereits adressiert worden, wenn auch seltener. So erstellen Lebiere und Anderson (2001) ein einfaches ACT-R Modell, mit dem es möglich ist Multi-Tasking und Cognitive Workload (mentale Beanspruchung) in einer en-route-Aufgabe zu beschreiben. Sie gehen bei der Analyse mentaler Beanspruchung allerdings weniger theoriegetrieben vor. Da in ACT-R kein Konzept von mentaler Beanspruchung enthalten ist, beschrieben Lebiere und Anderson (2001) die mentale Beanspruchung lediglich als skaliertes Verhältnis zwischen der Zeit, die das Modell mit kritischen Teilaufgaben (critical unit tasks) verbringt und der Gesamtzeit, in der es Aufgaben verrichtet. Kritische Teilaufgaben sind hierbei einerseits die Ausführung von Aktionen und andererseits die Suche nach neuen Ereignissen, die eines Eingreifens bedürfen. Neben der eher losen theoretischen Fundierung mentaler Beanspruchung sind auch die verwendete Aufgabe und die damit verbundene Simulation als sehr artifiziell zu betrachten. Es ist also anzunehmen, dass eine Generalisierung der Ergebnisse und eine Anwendung des Modells auf realistische Aufgaben nicht ohne weiteres möglich sind.

Neben dem Modell von Lebiere und Anderson (2001) fanden in der Flugverkehrskontrolle zahlreiche Versuche statt, die mentale Beanspruchung von Lotsen mit Hilfe mathematisch ermittelter Indizes anhand objektiv beobachtbarer Variablen zu berechnen (z.B. TLI; vgl. Kapitel 2.3). Dieser Ansatz birgt eine relative Einfachheit in sich, ist allerdings weder kognitiv plausibel noch ganzheitlich, da nur objektive Komplexitätsfaktoren (wie beispielsweise die Anzahl der zu kontrollierenden Flugzeuge) in die Berechnung solcher Indizes einfließen. Dass diese allerdings nur einen Teil der Einflussfaktoren mentaler Beanspruchung darstellen, ist bereits in Kapitel 2.3 dargestellt worden. Es gilt demnach, die hier offensichtlich werdenden Lücken zu schließen und ein ganzheitliches Modell zu erstellen, welches sowohl quantitative Abschätzungen mentaler

[7] Die abstrakte Darstellung der Aufgabe ist höchstwahrscheinlich der Ende der 80er, Anfang der 90er Jahre zur Verfügung stehenden begrenzten Rechnerkapazität geschuldet.

Beanspruchung von Towerlotsen, sowie Ganzheitlichkeit als auch kognitive Plausibilität und zu einem gewissen Grade Validität erreicht.

3.4 Kognitive Modellierung Mentaler Beanspruchung

Nachdem sowohl im Kapitel 2 als auch im Kapitel 3 Lücken hinsichtlich der Erfassung mentaler Beanspruchung (keine Vorhersage mentaler Beanspruchung möglich), sowie deren modellhafter Beschreibung (unzureichende theoretische Fundierung, Fokussierung auf den en-route Bereich) sichtbar gemacht wurden, gilt es nun, diese Lücken zu schließen. Zunächst kann festgestellt werden, dass die zur Verfügung stehenden Methoden zur Erfassung mentaler Beanspruchung keine hinreichende Güte besitzen, um im Feld als zuverlässiger Indikator mentaler Beanspruchung von Towerlotsen eingesetzt werden zu können, oder als Grundlage für die Entwicklung adaptiver Automatisierungs- und Unterstützungssysteme zu dienen.

Einerseits sind Diagnostizität und Sensitivität z.B. der Herzrate zu gering, um eine verlässliche Erfassung mentaler Beanspruchung zu erhalten. Andererseits ist die zeitliche Auflösung subjektiver Maße zu gering und die Interferenz mit der Hauptaufgabe von Leistungsmaßen zu groß. Es ist also nötig, eine Möglichkeit zu finden, wie diese Nachteile ausgeglichen werden können. Die Lösung kann hier in der kognitiven Modellierung und Simulation liegen. Für die Entwicklung eines kognitiven Modells werden zu Beginn zwar Ressourcen wie Zeit und Geld, sowie Maße zur Validierung benötigt, allerdings kann ein einmal aufgestelltes Modell immer wieder genutzt werden, um Prognosen hinsichtlich zukünftiger Beanspruchung von Lotsen zu machen um beispielsweise eine bessere Verkehrsplanung zu ermöglichen. Ferner kann ein solches Modell verwendet werden, um neuartige Unterstützungssysteme hinsichtlich ihres Nutzens überprüfen zu können, ohne dass Geld in die Entwicklung von Prototypen und aufwändige Feldstudien und Messungen investiert werden muss. Ein fortwährender Einsatz dieses Modells, kann daher aufwändige Messungen im Feld und die damit verbundenen Beeinträchtigungen der Lotsen überflüssig machen, sowie die Entwicklung einschlägiger Unterstützungssysteme vorantreiben, wobei Zeit und Geld gespart werden kann. Es ist daher sinnvoll und notwendig, ein lauffähiges Modell mentaler Beanspruchung von Towerlotsen zu erstellen, um Einschätzungen darüber zu erhalten, wie ein Lotse während eines bestimmten Verkehrsszenarios belastet ist bzw. sein wird.

Zur Validierung des Modells kann eine Kombination der vorgestellten Maße mentaler Beanspruchung in einer simulierten Umgebung verwendet werden. Eine solche Validierung hat daher auch in einer etwas abstrakteren Umgebung stattzufinden (mikrokognitiv), um eine Beeinträchtigung von Lotsen am Arbeitsplatz zu vermeiden und die Folgen von Fehlhandlungen gering zu halten. Bei der Verwendung eines simulierten Arbeitsplatzes, sollte allerdings dennoch

auf eine hinreichend große Korrespondenz mit einer realen Arbeitsumgebung geachtet werden (makrokognitiv).

Die Umsetzung des Modells kann anhand der beschriebenen kognitionspsychologischen Theorien erfolgen (vgl. Kapitel 2). Es konnte aus verschiedenen Theorien herausgearbeitet werden, dass vor allem Begrenzungen des Arbeitsgedächtnisses die Entstehung mentaler Beanspruchung bedingen. Somit sollte auch die mentale Beanspruchung über die Belastung des Arbeitsgedächtnisses, konkret über die Anzahl der Chunks, die zu einem Zeitpunkt im Arbeitsgedächtnis gespeichert sind, modellhaft abbildbar sein (mikrokognitiv). Aufgrund der noch sehr unklaren bzw. teilweise widersprüchlichen Befunde zu Ressourcengrenzen der Multi-Komponenten-Modelle des Arbeitsgedächtnisses, ist die Anwendung einer Single-Limited Capacity Theory mit einer maximalen Arbeitsgedächtniskapazität von zehn speicherbaren Chunks zu bevorzugen, wenn von Domainexperten ausgegangen wird (vgl. Bainbridge, 1975; Bisseret, 1970, 1971; Sperandio, 1969). Ebenso kann davon ausgegangen werden, dass diese maximale Anzahl von Chunks bei multiplen Ressourcenmodellen auf eine bestimmte (noch unbekannte) Weise auf die vorhandenen Sub-Systeme aufgeteilt ist. Es ist daher sinnvoll, das Arbeitsgedächtnis vereinfacht zunächst als durch eine einzelne Kapazität limitiert zu betrachten. Hierbei wird die Ansicht vertreten, dass der Operateur zur Aufgabenausführung Informationen im Arbeitsgedächtnis aufrechterhalten muss (Baddeley & Hitch, 1974; Daneman & Carpenter, 1980). Diese Ansicht steht im Gegensatz zu Annahmen des Arbeitsgedächtnisses in Theorien der Produktionensysteme (z.B. Newell, 1990). In ACT-R beispielsweise sind beide Ansätze implementiert (Anderson, Reder, & Lebiere, 1996), wobei allerdings keine explizite Arbeitsgedächtnisstruktur existiert. Vielmehr sind in Anlehnung an die Theorien der Produktionensysteme die aktuell aktivierten Informationen des deklarativen Gedächtnisses, auf die Produktionsregeln angewendet werden, als eine Art Arbeitsgedächtnis zu bezeichnen. Es wird dabei, wie auch bei Cowan (1999, 2010), als aktuell aktivierter Teil des Langzeitgedächtnisses gesehen. Eine Kapazitätsgrenze im weitesten Sinne liegt hierbei in der Begrenzung der gesamten Grundaktivierung. Diese weist allerdings ebenfalls Ähnlichkeit mit dem Konzept einer einzelnen Kapazitätsbegrenzung auf, wie sie beispielsweise auch Kahnemann (1973) beschreibt (Anderson et al., 1996). Die Ansicht, dass Informationen aufrechterhalten werden müssen, wird in ACT-R auf prozedurales Wissen angewendet. Hierbei liegt die Begrenzung in den einzelnen Puffern, in denen immer nur ein Element gleichzeitig aufrechterhalten werden kann.

Darüber hinaus ist aber auch die ganzheitliche Betrachtung des Mensch-Maschine-Systems Flugverkehrskontrolle nötig, so dass auch Flughafenprozess und Lotsen-Flughafen-Interaktionen, sowie Prozesse der Wahrnehmung und Handlungsausführung des Lotsen mit beschrieben werden müssen. Dies ermöglicht die Betrachtung der Auswirkungen kognitiver

Komponenten des Lotsen auf die Leistung des Gesamtsystems, benötigt allerdings eine eher abstraktere Betrachtung kognitiver Funktionen (makrokognitiv; vgl. Smieszek et al., 2013). Dies birgt darüber hinaus die Möglichkeit der Wiederverwendbarkeit des Modells und es einem bereiten Anwendungsspektrum zuzuführen. Für die Entwicklung des Modells sind daher zunächst Analysen des Systems und der Aufgaben von Lotsen im Feld durchzuführen (makrokognitiv). Für die genauere Beschreibung einzelner kognitiver Funktionen ist ein downscaling unter Verwendung kognitionspsychologischer Theorien, die dieses Phänomen beschreiben, notwendig (mikrokognitiv; vgl. Kapitel 3.2.3).

3.5 Kapitelzusammenfassung

Es wurde zunächst beschrieben, wie ein Mensch-Maschine-System definiert ist, und wie die Modellierung im Allgemeinen und die kognitive Modellierung und Simulation im Speziellen zur Gestaltung und Analyse von Mensch-Maschine-Systemen beitragen kann (Kapitel 3.1). Hierbei wurde die Interdisziplinarität der kognitiven Modellierung herausgestellt, wodurch sich aber auch verschiedenartige Herangehensweisen zur modellhaften Beschreibung von Mensch-Maschine-Systemen ergeben. Die kognitionswissenschaftliche Herangehensweise (mikrokognitive Modellierung) zeichnet sich dadurch aus, dass die genaue Funktionsweise kognitiver Prozesse im Mittelpunkt der Modellbildung steht und die Untersuchung von Problemstellungen eher grundlagenwissenschaftlich erfolgt. Modelle werden häufig auf Basis sogenannter kognitiver Architekturen erstellt (vgl. Kapitel 3.2.1).

Die ingenieurwissenschaftliche Herangehensweise (makrokognitive Modellierung) hingegen ist eher anwendungsorientiert und weniger auf die genaue Funktionsweise kognitiver Prozesse fixiert, sondern vielmehr auf den Beitrag kognitiver Prozesse zur Gesamtleistung eines Mensch-Maschine-Systems (vgl. Kapitel 3.2.2). Es wird allerdings argumentiert, dass sich die beiden Herangehensweisen nicht gegenseitig ausschließen, sondern sie vielmehr zwei Punkte eines Kontinuums darstellen und eine Verbindung beider möglich und sogar erstrebenswert ist. Die vorliegende Arbeit lässt sich dabei eher den makrokognitiven Ansätzen zuordnen, stellt allerdings die Verbindung zu mikrokognitiven Ansätzen her (vgl. Kapitel 3.2.3). Teilweise wurden bereits durch existierende mikro- und makrokognitive Modelle verschiedene ähnliche Problemstellungen, wie sie hier vorliegen, adressiert, wovon einige ausgewählte Modelle bezüglich Fragestellungen in der Flugverkehrskontrolle, der Luftfahrt, sowie zu mentaler Beanspruchung vorgestellt wurden. Es ist festzuhalten, dass einige Ansätze dabei vor allem auf physikalisch-technische Problemstellungen in der Flugverkehrskontrolle fokussieren, menschliches Verhalten allerdings nicht speziell adressieren. Ferner existieren formale Modelle, welche zur Analyse menschlichen Verhaltens und im speziellen menschlicher Entscheidungsheuristiken in der Flugverkehrskontrolle dienen, wobei auch hier kognitive

Prozesse eher wenig detailliert betrachtet werden. Mikrokognitive Ansätze beschäftigen sich vor allem mit der Untersuchung von Fragestellungen der Entscheidungsfindung, des Fertigkeitserwerbs und mentaler Beanspruchung und berücksichtigen kognitive Prozesse mit hinreichender Genauigkeit, fokussieren allerdings meist auf den en-route Bereich nur vereinzelt auf Aufgaben, die mit denen von Towerlotsen vergleichbar sind. Ferner werden der Kontext und die Aufgabe, meist nur sehr vereinfacht abgebildet. Speziell in den vorliegenden Modellen mentaler Beanspruchung ist keine theoretische Fundierung vorgenommen worden, so dass hier Handlungsbedarf identifiziert wurde (vgl. Kapitel 3.3).

Durch die kognitive Modellierung und Simulation kann ein Hilfsmittel geschaffen werden, welches Nachteile vorhandener Methoden ausgleichen und neue Möglichkeiten eröffnen kann: So können beispielsweise Vorhersagen der mentalen Beanspruchung unter veränderten Bedingungen gemacht werden (z.B. veränderte Flugpläne, neue Unterstützungssysteme), was derzeit nur mittels groß angelegter Simulatorstudien oder Testläufen vor Ort und entsprechender Erhebung subjektiver, physiologischer und Leistungsmaße möglich ist. Zwar ist die Erstellung kognitiver Modelle und Simulationen mit hohem Aufwand und Kosten verbunden, jedoch ist der Nutzen durch die wiederholte Anwendung um ein vielfaches größer (vgl. Kapitel 3.4). Ein solches kognitives Modell bzw. eine kognitive Simulation wird auf Basis einschlägiger Felduntersuchungen und unter Verwendung kognitionspsychologischer Theorien der Informationsverarbeitung in dieser Arbeit erstellt werden.

4 Entwicklung eines makrokognitiven Modells der Flughafen-verkehrskontrolle (MATriCS)

Im Folgenden sollen die theoretischen Vorarbeiten zur Entwicklung des Modells anhand der Richtlinien für die Entwicklung kognitiver Simulationen, die Cacciabue (1998) vorschlägt, beschrieben werden. Hierzu wird zunächst allgemein auf die einzuhaltenden Schritte bei der Entwicklung kognitiver Simulationen eingegangen (Kapitel 4.1). Die nachfolgenden Kapitel beschreiben jeweils im Detail die durchgeführten Arbeiten innerhalb der einzelnen Prozessschritte. Es werden zunächst die Problemgrenzen definiert (Kapitel 4.2), anschließend die verwendeten Aufgabenanalysen vorgestellt (Kapitel 4.3), sowie auf die Analyse der Arbeitsumgebung eingegangen (Kapitel 4.4). Die Kapitel 4.5 und 4.6 beschreiben die Erstellung eines theoretischen Modells und die Auswahl des Beschreibungsmittels. Kapitel 4.7 fasst das Kapitel zusammen.

4.1 Vorgehensweise zur Entwicklung Kognitiver Simulationen

Cacciabue (1998b) beschreibt ein allgemeines Ablaufschema, welches als Richtlinie der durchzuführenden Analyseschritte für die Entwicklung kognitiver Simulationen eingehalten werden sollte. Er definiert sechs Schritte in einem iterativen Prozess, die durchzuführen sind, um eine solide und zuverlässige Simulation zu erhalten. Diese sind in Abbildung 5 schematisch dargestellt. Als ersten und wichtigsten Schritt beschreibt Cacciabue (1998b) die *Definition der Problemgrenzen und des Simulationszieles.* Hier sollte einerseits das vorliegende Problem, sowie die Ziele und das erwartete Ergebnis der Simulation bestimmt werden. Unter Anderem ist hier auch der Detailgrad festzulegen, den das zu untersuchende Problem erfordert (Cacciabue, 1998b). Als zweiten und dritten Schritt sieht Cacciabue (1998b) die Durchführung einer *Kognitiven Aufgabenanalyse*, sowie einer *Feldstudie der Arbeitsumgebung* vor. Beide können entweder unabhängig voneinander oder in enger Verknüpfung stattfinden. Beide Schritte sind dabei von großer Wichtigkeit für den Entwicklungsprozess der kognitiven Simulation. Die *Auswahl des theoretischen Modells* erfolgt als vierter Schritt. Dieser Schritt ist eng mit dem Detailgrad der Simulation und dem untersuchten Arbeitsumfeld verbunden. Dabei stellt die Auswahl eines theoretischen Modells menschlichen Verhaltens und kognitiver Prozesse nur die Grundlage dar, die weiter unter Anwendung des Beschreibungsmittels in das Arbeitsumfeld integriert werden muss. Wie in Abbildung 5 zu erkennen ist, können nach Cacciabue (1998b) zwei Ziele verfolgt werden: *Deskriptive* oder *numerische Simulation*. Für die Entwicklung einer deskriptiven Simulation, müssen nicht alle sechs Schritte befolgt werden, da die zur numerischen Simulation nötige Implementierung in einer Programmierumgebung entfällt (Cacciabue, 1998b). Für die hier angestrebte numerische Simulation ist allerdings im Schritt fünf

zunächst die *Auswahl eines geeigneten Beschreibungsmittels* zu treffen[8], welche den Modellierer in die Lage versetzt, durch die *Implementierung in einer Programmierumgebung* seine gestellten Ziele zu erreichen.

Wie ebenfalls in Abbildung 5 zu erkennen ist, sind zahlreiche Iterationen zwischen den einzelnen Schritten vorgesehen, falls in einem bestimmten Schritt der Simulationsentwicklung die Notwendigkeit für Anpassungen identifiziert wird, so dass ein vorausgehender Schritt wiederholt bzw. vertieft werden muss.

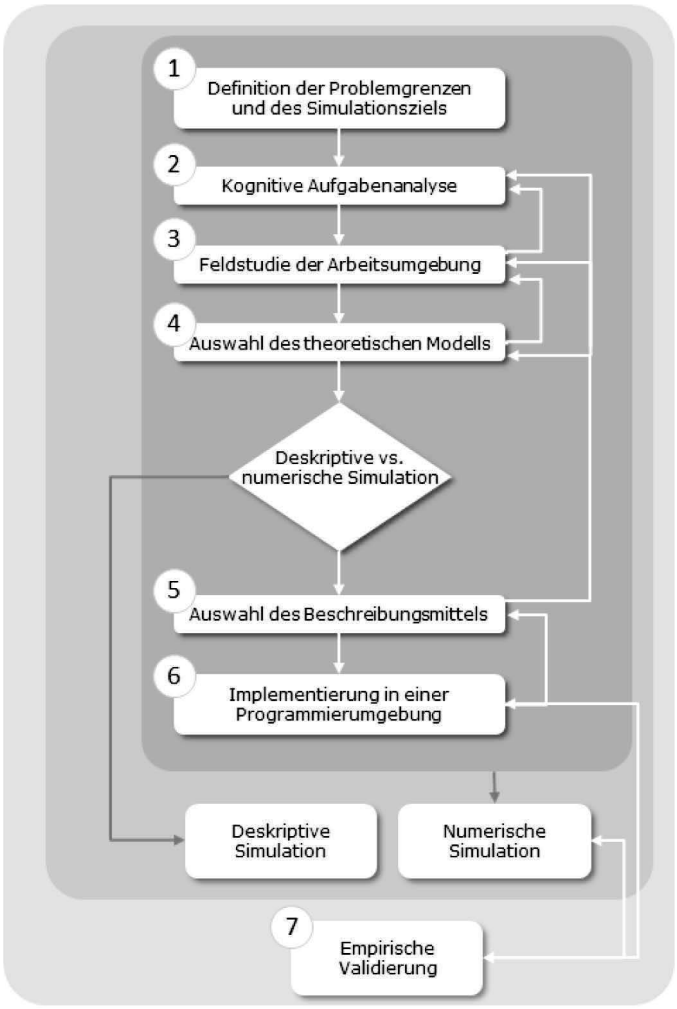

Abbildung 5: Richtlinien für die Entwicklung kognitiver Simulationen modifiziert nach Cacciabue (1998b) (eigene Darstellung)

Es ist außerdem darauf hinzuweisen, dass das dargestellte Flussschema nach Cacciabue (1998b) in einer eher ingenieurwissenschaftlichen Tradition der Modellbildung steht und daher einen anderen Fokus setzt als es in der Kognitionswissenschaft oder der Psychologie der Fall ist. Wie bereits im Abschnitt 3.2 beschrieben, sind Modelle in den Ingenieurwissenschaften oft nur „so genau wie nötig“ (Engell, 1992). Dies kann mit der makrokognitiven Sichtweise verglichen

[8] Der besseren Verständlichkeit wegen wurde die ursprüngliche Bezeichnung des fünften Schrittes „Auswahl des numerischen Algorithmus“ durch die Bezeichnung „Auswahl des Beschreibungsmittels“ ersetzt.

werden. In der Psychologie und den Kognitionswissenschaften zielt die Modellbildung eher auf eine möglichst große Übereinstimmung zwischen Original und Modell ab (Jürgensohn, 2002), was der mikrokognitiven Sichtweise entspricht.

Gleichzeitig lässt das Cacciabue-Framework einen Schritt vermissen, der vor allem in der psychologischen Modellbildung, aber auch in den Ingenieurwissenschaften als essentiell anzusehen ist: Die Validierung, also der Vergleich von Modellverhalten mit experimentellen Daten. Durch die bloße Erstellung einer kognitiven Simulation kann nicht per se davon ausgegangen werden, dass diese auch etwas über den Untersuchungsgegenstand aussagt. Vielmehr ist es die Aufgabe des Modellierers, über die Validierung den Bezug zwischen dem Modell und dem realen Untersuchungsgegenstand herzustellen (Möhlenbrink, 2011; vgl. auch Bub & Lugner, 1992).

In der Psychologie geschieht eine solche Validierung mit sehr ausgefeilten Methoden der statistischen Validitätsmessung. In den Ingenieurwissenschaften sind die eingesetzten Techniken meist sehr viel einfacher, und geschehen manchmal sogar nur anhand des gesunden Menschenverstandes des Modellierers oder Modelliererteams durch heuristische Evaluation oder „mental Simulation“ (Jürgensohn, 2002). Dies liegt vor allem an der Tatsache, dass ingenieurwissenschaftliche Modelle meist sehr viel mehr freie Parameter enthalten als psychologische Modelle. Modelle der Psychologie und damit verbundene empirische Untersuchungen versuchen immer durch die Fokussierung auf eine oder einige wenige Variablen und die Eliminierung möglichst aller Störvariablen die Komplexität so gut wie möglich einzuschränken. Dies bringt den Vorteil mit sich, Informationen zu erhalten, die genau den Sachverhalt betreffen, der im Mittelpunkt der Untersuchung steht. Der Nachteil ist allerdings, dass die Aufklärungsgüte sich eben auch nur auf diesen einen Parameter beschränkt und die externe Validität damit nicht immer gegeben ist, da in der Realität immer viele Parameter miteinander interagieren und sich gegenseitig beeinflussen. Auf der anderen Seite stehen Modelle mit vielen freien Parametern sehr stark in der Kritik, da ein Modell nur genügend freie Parameter enthalten muss, um in der Lage zu sein, auf jeden beliebigen Datensatz mit hoher Präzision zu passen (Roberts & Pashler, 2000; Schunn & Wallach, 2005). Solche „Stellschraubenmodelle“ können einen bestimmten Sachverhalt oder einen bestimmen Datensatz sehr gut abbilden. Eine Generalisierung auf andere Datensätze ist allerdings ohne Anpassung der „Stellschrauben“ nicht möglich. Das bedeutet, dass die freien Parameter neu gesetzt werden müssen, um auf einen anderen Datensatz zu passen, was zum sogenannten Overfitting[9] führt (Myung, 2000). Modelle mit wenigen freien Parametern passen dagegen auf

[9] Mit Overfitting (Überanpassung) wird die Anpassung eines Modells an einen vorliegenden Datensatz bezeichnet. Ebenso kann Overfitting als Spezifizierung eines Modells beschrieben werden, welches zu viele erklärende Variablen enthält.

den Einzelfall schlechter, sind aber in der Lage mit konstanter Güte eine große Menge unterschiedlicher Fälle zu beschreiben, ohne dass Anpassungen vorgenommen werden müssen.

Prinzipiell ist eine Validierung im Sinne einer empirischen, psychologischen Testung somit nur dann möglich, wenn bereits bei der Problemstellung der Fokus auf einen ganz bestimmten Aspekt des menschlichen Verhaltens gelegt wird, der im Labor sehr gut zu extrahieren ist. Bei Mensch-Maschine-Modellen, denen ein eher ingenieurwissenschaftlicher Ansatz zugrunde liegt (also ganzheitlich bzw. makrokognitiv), sind allerdings meist mehrere bis sehr viele verschiedene Faktoren Teil der Betrachtung, die nicht immer experimentell zu kontrollieren sind und ihrerseits miteinander interagieren. Für eine Modellvalidierung im Bereich der makrokognitiven Modellierung bringt dies somit die Schwierigkeit mit sich, dass zwei Dinge vereint werden müssen, die eigentlich nicht vereinbar sind, nämlich experimentelle Kontrolle weniger Variablen auf der einen Seite und die Berücksichtigung und Vorhersage vieler Variablen auf der anderen. Es ist daher ein Balanceakt notwendig, der einerseits der ingenieurwissenschaftlichen Tradition durch möglichst ganzheitliche Beschreibung des Mensch-Maschine-Systems Rechnung trägt und makrokognitiv erfolgt, und der gegebenen Problemstellung aber hinreichend Genüge tut. Andererseits soll auch die psychologische Tradition der empirischen Modellvalidierung mittels experimenteller Kontrolle weniger Variablen (also mikrokognitiv) als notwendiger Schritt des Modellbildungsprozesses Anerkennung finden. Durch diese Synthese beider Wissenschaftsbereiche kann eine neue Perspektive und eine gute Ergänzung zur klassischen kognitiven Modellierung geschaffen werden und gleichzeitig der ingenieurwissenschaftliche „gesunde Menschenverstand“ durch fundierte wissenschaftliche Techniken ergänzt werden. Ein solcher Balanceakt kann dann gelingen, wenn eine schrittweise Validierung des Modells durchgeführt wird, in dem jede betrachtete Variable für sich untersucht und ihr Einfluss experimentell bestimmt wird, während alle anderen Variablen konstant gehalten werden. Bereits bei der Definition der Problemgrenzen ist daher festzulegen, welche Variablen berücksichtigt werden sollen und wie das Modell im Anschluss validiert werden kann.

4.2 Definition der Problemgrenzen

Der erste und wichtigste Schritt zur Entwicklung kognitiver Simulationen besteht in der genauen Definition des vorliegenden Problems, der Ziele und der erwarteten Ergebnisse der Simulation (Cacciabue, 1998b). Für die vorliegende Arbeit wurden einige der Probleme bereits in den vorangegangenen Kapiteln 1-3 herausgearbeitet, so beispielsweise das steigende Flugverkehrsaufkommen und die damit verbundene Bedeutung der Erfassung mentaler Beanspruchung (vgl. Kapitel 1 und 2), sowie das Fehlen adäquater kognitiver Modelle und Simulationen mentaler Beanspruchung im Bereich der Flughafenverkehrskontrolle.

Es steht die Domäne Flugverkehrskontrolle an Flughäfen im allgemeinen Fokus. Hierbei soll grundlegend eine makrokognitive Herangehensweise verfolgt werden: Das System der Flugverkehrskontrolle an Flughäfen soll möglichst ganzheitlich beschrieben und eine Vielzahl an Variablen prinzipiell im Modell abgebildet werden. Hierzu gehört es sowohl den *zu kontrollierenden Prozess (1)* als auch die *menschliche Kognition (2)*, sowie *deren Interaktionen (3)* modellhaft zu beschreiben (Cacciabue, 1998a, 1998b).

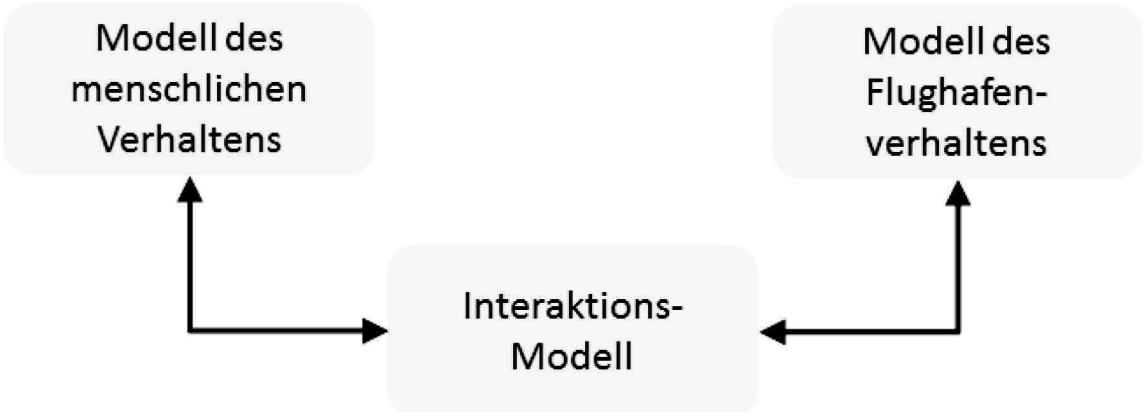

Abbildung 6: Dreiteilung von Mensch-Maschine-Modellen (eigene Darstellung nach Cacciabue, 1998a)

Der zu *kontrollierende Prozess* ist hierbei der Flughafen und die darauf stattfindenden Verkehrsbewegungen. Dabei sind sowohl physikalische Gesetzmäßigkeiten, die trivial erscheinen mögen (z.B. dass Zurücklegen von Entfernungen in Abhängigkeit von der Geschwindigkeit), als auch regulatorische, vom Menschen geschaffene, Begrenzungen und Einschränkungen (z.B. die maximal erlaubte Rollgeschwindigkeit von Flugzeugen auf den Rollwegen) zu berücksichtigen. Diese Variablen sind veränderlich in der Zeit und erzeugen so dynamische Veränderung des Flughafens. Sie sind durch den Lotsen zu kontrollieren, welcher durch seine *kognitiven Prozesse* in die Lage versetzt wird, die ihm gestellten Aufgaben zu bewältigen. Die Performanz des Lotsen ist dabei von seinen kognitiven Prozessen und Limitationen abhängig, welche ebenfalls durch eine Vielzahl Variablen beschrieben werden können (z.B. Motivation, Erfahrung, Arbeitsgedächtniskapazität usw.; vgl. Kap. 2.3). Durch die *Interaktion* mit dem Flughafenprozess, kann der Lotse Informationen über den Prozess abrufen und regulierend eingreifen. Die Zeitabhängigkeit der Interaktionen fügt dabei ein weiteres Element von Komplexität hinzu (Cacciabue, 1998a). Diese Interaktionen beschreiben den Austausch von Informationen zwischen Flughafenmodell und Lotsenmodell unter Verwendung verschiedener digitaler oder analoger Hilfsmittel, welche ihrerseits bestimmte Eigenschaften aufweisen (z.B. Güte der Funkübertragung, Verwendungsdauer des Funkkanales, Sichtbedingungen usw.). Hierbei laufen Flughafenprozess und kognitive Prozesse des Lotsen unabhängig voneinander ab, und synchronisieren sich nur zeitweise durch den Austausch von Informationen über das Interaktionsmodell. Diesen Zusammenhang stellt Abbildung 6 schematisch dar. Es können mit diesem Ansatz somit eine Vielzahl von Variablen abgebildet

werden, welche alle einen Einfluss auf die Performanz des Lotsen und damit die Performanz des Gesamtsystems besitzen.

Aufbauend auf den Arbeiten von Koros et al. (2003) sollen zur Eingrenzung des Variablenraumes für das vorliegende Modellierungsproblem diejenigen Faktoren Berücksichtigung finden, die Lotsen als wichtigste Einflussfaktoren bewertet haben. Hierbei zeigte sich, dass „hohes Verkehrsaufkommen" als wichtigster Faktor bewertet wurde. Ebenso sind „Kreuzen der aktiven Runway", „Überlastungen der Frequenz", „Unterschiedliche Leistungscharakteristika von Flugzeugen", „Einschränkungen der RWY[10] oder TWYs[11]", „Konfiguration der RWY/TWYs", „Maßnahmen der Verkehrsplanung", „Wirbelschleppen[12]", „Ausbildung am Arbeitsplatz", „schlechte Sichtbedingungen (durch Wetter)", „schlechtes Wetter", „unvertraute Piloten", „Verkehr von Fahrzeugen", und „schlechtes Englisch von Piloten" oft genannte Komplexitätsfaktoren. Einige dieser Faktoren sind allerdings mit einem modellgetrieben Ansatz kaum oder schlecht beschreibbar, wie beispielsweise „Maßnahmen der Verkehrsplanung" oder „Ausbildung am Arbeitsplatz". Viele der Faktoren, die als wichtig erachtet wurden, können modelliert werden und sollen in der vorliegenden Arbeit Beachtung finden: „hohes Verkehrsaufkommen", „Überlastungen der Frequenz", „Wirbelschleppen", „schlechte Sichtbedingungen", „Verkehr von Fahrzeugen". Wie die einzelnen Faktoren im Detail im Modell Berücksichtigung finden, wird im Kapitel 5 erläutert.

Um den Problemraum sinnvoll einzugrenzen und damit gleichzeitig die Möglichkeit für eine experimentelle Modellvalidierung zu schaffen, soll das Modell im vorliegenden Fall verwendet werden, um den Einfluss des Flugverkehrsaufkommens (der Anzahl der zu kontrollierenden Flugzeuge innerhalb eines bestimmten Verkehrsszenarios) auf die mentale Beanspruchung des Lotsen (hier operationalisiert durch die Anzahl der im Arbeitsgedächtnis gespeicherten Chunks; vgl. Kapitel 3.4) abzubilden. Die maximale Anzahl an speicherbaren Chunks wird dabei auf Grundlage zahlreicher Studien auf zehn festgelegt (Bainbridge, 1975; Bisseret, 1971; Sperandio, 1969). Alle anderen „äußeren" Variablen werden im Modell zwar vorgesehen, werden aber soweit möglich konstant gehalten. Um das Modell möglichst generisch zu halten, wird sichergestellt, dass eine Abbildung verschiedener Flughäfen mit unterschiedlichster Konfiguration relativ leicht in das Modell einzubinden ist. Zusätzlich ermöglicht diese Vorgehensweise die Konfiguration des Flughafens als Komplexitätsfaktor (vgl. Koros et al. 2003)

[10] Runway: Start-/Landebahn
[11] Taxiway: Rollwege
[12] Das Starten und Landen von Flugzeugen erzeugt enorme Luftverwirbelungen, welche das Landen und/oder Starten direkt nachfolgender Flugzeuge erschweren bis unmöglich machen würden. Daher sind adäquate Abstände, sogenannte Staffelungsminima, zwischen zwei landenden bzw. startenden Flugzeugen einzuhalten.

mit berücksichtigen zu können[13]. Um den Einfluss der Anzahl der innerhalb eines bestimmten Zeitraumes zu bearbeitenden Flugzeuge auf die mentale Beanspruchung untersuchen zu können, ist es zusätzlich von Bedeutung, dass beliebige Verkehrsszenarien auf dem Flughafen mit hinreichender Genauigkeit beschrieben werden können.

Das *Lotsenmodell* wird anhand der Phasen der Handlungsregulation (Hacker, vgl. Kapitel 4.5) modularisiert. Die für die vorliegende Problemstellung im Mittelpunkt stehenden Funktionen sind mit ausreichender Gründlichkeit abzubilden. Der ganzheitliche Anspruch des Modells verlangt aber gleichzeitig die (zumindest funktionelle) Abbildung der nicht im Fokus der Arbeit stehenden Prozesse, sofern sie essentiell für die Erfüllung der Aufgabe des Lotsen sind. Ferner wird festgelegt, dass nur ein Lotse den gesamten Ablauf eines Flughafens kontrolliert und die Arbeit somit nicht, wie in der Realität üblich, auf zwei Positionen aufgeteilt wird (vgl. Kapitel 4.4). Der modulare Aufbau ermöglicht allerdings Anpassungen und/oder Ergänzungen, um zur Modellierung anderer Problemstellungen, beispielsweise die Einbindung eines zweiten Lotsen- oder Flughafenmodells zu gewährleisten. Zur Interaktion zwischen Lotsen- und Flughafen-Modell sollen ferner verschiedene vom Lotsen verwendete Hilfsmittel abgebildet werden (vgl. Kapitel 4.4). Als weitere Vereinfachung wird festgelegt, dass Piloten, im Gegensatz zum Lotsen, als zweite menschliche „Komponente" des Systems Flughafen, nicht explizit mit modelliert werden, sondern als Teil des Flughafenmodells angesehen werden.

Die wichtigsten Annahmen und Vereinfachungen sind im Folgenden nochmals in Kürze zusammengefasst:

- Es wird eine ganzheitliche Beschreibung der Domäne Flughafenverkehrskontrolle angestrebt.
- Es erfolgt eine Dreiteilung des Modells in Flughafen-, Interaktions- und Lotsenmodell
- Das *Flughafenmodell* beschreibt physikalische Prozesse (z.B. Flugbewegungen) sowie regulatorische Einschränkungen (z.B. zulässige Rollgeschwindigkeiten) des zu kontrollierenden Systems.
- Das *Interaktionsmodell* beschreibt den Informationsaustausch zwischen Flughafen- und Lotsenmodell. Es werden die wichtigsten Informationsquellen (Flugstreifen, Außensicht, Radar und Funkkommunikation) modelliert.
- Ferner werden die wichtigsten Komplexitätsfaktoren modelliert: Verkehrsaufkommen, Belastung der Funkfrequenz, Wirbelschleppen, schlechte Sichtbedingungen, Verkehr von Fahrzeugen, Flughafenlayout.

[13] Die Flughafenkonfiguration wurde bei Koros et al. (2003) als „Airspace Configuration" („Luftraum-konfiguration") bezeichnet.

- Zur sinnvollen Eingrenzung des Problemraumes wird nur das Verkehrsaufkommen als wichtigster Komplexitätsfaktor betrachtet. Alle anderen Komplexitätsfaktoren werden konstant gehalten.
- Das *Lotsenmodell* wird anhand der Phasen der Handlungsregulation (Hacker, 1986 vgl. Kapitel 4.5) modularisiert.
- Phasen, die nicht dem Fokus der Problemstellung der Arbeit stehen, werden zumindest funktionell abgebildet, sofern sie für die Aufgabenausführung des Lotsen essentiell sind.
- Zur Modellierung dieser Phasen werden weitere kognitive Theorien implementiert (beispielsweise Wickens (1984) Theorie multipler Ressourcen zur Beschreibung der Informationsaufnahme).
- Zur quantitativen Beschreibung der mentalen Beanspruchung des Lotsenmodells wird ferner ein Arbeitsgedächtnismodul implementiert, welches sich an Single-Limited Capacity Theorien orientiert und eine maximale Kapazität von zehn Luftfahrtspezifischen Chunks besitzt (vgl. Kapitel 2.2 und 3.4)
- Ein Chunk enthält dabei jeweils die Information über ein Flugzeug
- Es wird festgelegt, dass nur ein Lotse (nur ein Lotsenmodell) den gesamten Ablauf eines Flughafens kontrolliert. Es wird keine Aufteilung auf zwei Lotsen (zwei Lotsenmodelle) vorgenommen (vgl. Kapitel 4.3 und 4.4)
- Piloten werden als zweite menschliche „Komponente“ des Systems nicht explizit modelliert, sondern als Teil des Flughafenmodells angesehen.

4.3 Aufgabenanalyse

Für die Erstellung des Lotsen- sowie des Flughafenmodells ist eine genaue Kenntnis der Abläufe eines Flughafens und der Aufgaben der Lotsen notwendig. Zu diesem Zwecke wurden neben Regelwerken, wie der Betriebsanweisung für die Flugverkehrskontrolle (BA-FVK; DFS, 2006) und der ICAO4444 (2007)[14], die den Ablauf der Flugverkehrskontrolle beschreiben, bereits vorhandene Aufgabenanalysen aus dem Bereich der Flugplatzkontrolle ausgewertet und konsolidiert. Ziel war es, diejenigen Aufgabenanalysen herauszufinden, die die nötige Tiefe aufweisen, um ein möglichst umfassendes Bild der Aufgaben von Towerlotsen zu generieren.

Es existiert eine Vielzahl von Aufgabenanalysen im Bereich Flugverkehrskontrolle, mit unterschiedlichstem Fokus. Dabei können zunächst zwei Unterscheidungskriterien getroffen werden: Die analysierte Position (en-route vs. Tower) und die Art der Analyse (hierarchisch vs. kognitiv). Der Unterschied zwischen hierarchischer (hierarchical task analysis, HTA; vgl. z.B. Kirwan & Ainsworth, 1992) und kognitiver Aufgabenanalyse (cognitive task analysis, CTA; vgl.

[14] ICAO: International Civil Aviation Organization (Internationale Zivilluftfahrt-Organisation)

z.B. Crandall, Klein, & Hoffman, 2006) besteht darin, dass bei der CTA die „innere Wissensstruktur“ des Menschen, also das Wissen, bzw. die Informationen, die der Benutzer zur Durchführung einer Aufgabe hat oder haben muss, analysiert werden: *„CTA covers a range of approaches used for looking at mental (hence cognitive) internal events or knowledge structures*“ (Kirwan & Ainsworth, 1992, S. 392). Bei der HTA hingegen ist das physische von außen beobachtbare Verhalten von Interesse. Da für die Modellierung der Leistung von Operateuren sowohl physische als auch die kognitive Aufgaben von Belang sind, wurden hier beide Arten von Aufgabenanalysen ausgewertet. Beiden gemeinsam ist allerdings die Zerlegung komplexer Aufgaben in weniger komplexe Unteraufgaben mit Hilfe von sogenannten Dekompositionsgrafen.

Das zweite Kriterium ist die Position, welche analysiert wurde. Da Forschungen bezüglich der Flugverkehrskontrolle in den letzten Jahrzehnten vor allem auf den en-route Bereich fokussiert waren (also denjenigen Lotsen, welche die Staffelung der Flugzeuge im oberen Luftraum koordinieren), existieren die meisten Analysen für diese Position (vgl. Kapitel 1 und 3).

Während der Recherchearbeiten stellte es sich ferner als schwierig heraus, Zugang zu einigen Originaldokumenten von Aufgabenanalysen zu erhalten. Daher wurden ebenso bereits existierende Zusammenfassungen von Aufgabenanalysen in die Untersuchung mit einbezogen. Buck, Biemans, Hilburn und van Woerkom (1996) fügten beispielsweise die Analysen von Alexander, Alley, Ammerman, Hostetler und Jones (1988), Endsley und Rodgers (1994), ETACHIP (1996) und Jackson (1989), sowie die Analyse von Cox (1994a) zusammen und unterteilten die Aufgabe des Lotsen in diverse Unteraufgaben. Hierbei werden genauere Informationen auch nur für den en-route Bereich gegeben, obwohl Cox (1994a) sehr detailliert auf Towerlotsen eingeht. Den gleichen Fokus setzt die integrierte Aufgabenanalyse von Dittmann, Kallus und Van Damme (2000), die lediglich einen Vergleich zwischen der en-route und der Towerposition anstellt.

Eine Zusammenfassung verschiedener Aufgabenanalysen, die sich vor allem auf den Bereich Towerlotsen fokussiert, wurde im Jahr 2006 durch die EUROCONTROL[15] durchgeführt (Tavanti, 2006a). Dieses Dokument fasst vier große Aufgabenanalyse-Studien (drei von der EUROCONTROL selbst, eine von der FAA[16]) zusammen. Zum einen wird wiederum die Integrierte Studie von Kallus, Van Damme und Dittmann (1999) beschrieben, die allerdings aus der hier vorgestellten Betrachtung ausgeschlossen werden muss, da sie auf die en-route Position fokussiert. Zum anderen werden zwei hierarchische Aufgabenanalysen der EUROCONTROL beschrieben, welche als Grundlage für das Design neuer Systeme dienen und speziell auf die

[15] EUROCONTROL: European Organisation for the Safety of Air Navigation – Internationale Organisation zur Koordination in der Luftverkehrskontrolle in Europa
[16] FAA: Federal Aviation Administration – Bundesluftfahrbehörde der USA

Aufgaben von Towerlotsen eingehen – eine aus dem MANTEA[17] Projekt (Paul, Zografos, & Hesselink, 2000; Rossi et al., 1996), die andere aus dem ATHOS[18] Projekt (Courboulay & Kahn, 1996; Marti, 1998, 2000). Als viertes wird eine bereits 1988 durchgeführte hierarchische Aufgabenanalyse der FAA beschrieben, welche die wohl umfangreichste der hier beschriebenen Analysen darstellt (Ammerman, Becker, et al., 1988). Die Grundlage für diese Arbeit bilden damit die eben beschriebenen drei Aufgabenanalysen der Towerpositionen. Ferner wurde eine zusätzliche hierarchische Aufgabenanalyse hinzugezogen, die 2009 in Deutschland von der Firma Human-Factors-Consult für das DFS-Projekt VICTOR[19] durchgeführt wurde (Human-Factors-Consult, 2009), sowie eine Analyse des Royal Air Force Institute of Aviation Medicine (Cox, 1994a, 1994b), welche sich ebenfalls explizit mit Towerlotsen beschäftigen.

4.3.1 Aufgaben, Verantwortlichkeiten, Verpflichtungen

Zusätzlich zu konkreten Aufgaben ist in der Literatur oft auch von Verantwortlichkeiten und Verpflichtungen die Rede, welche allerdings klar von Aufgaben zu unterscheiden sind und eher mit den Bearbeitungsprioritäten von Verkehrspiloten (aviate, navigate, communicate) zu vergleichen sind und demzufolge wenig Details zur Lotsentätigkeiten beschreiben. Es handelt sich also mehr um höhergestellte Zielsetzungen, die mit konkret auszuführenden Aufgaben und Handlungen zu erreichen sind. So beschreiben Wickens, Mavor und McGee (1997) drei „key responsibilities" also drei Verantwortlichkeiten:

- Freigaben erteilen
- Bodenverkehr leiten
- Abgabe abfliegender und Annahme ankommender Flugzeuge vom Radarlotsen

Wittbrodt und Thüring (2010) beschreiben vier „Hauptaufgaben" von Towerlotsen:

- communication
- coordination
- monitoring/direct information search
- logging/documentation

Eine zusätzliche Pflicht des Lotsen sieht Nolan (2011) in der Verhinderung von Runway Incursions[20]. Diese Pflicht steckt bereits in der von der ICAO (ICAO4444, 2007) vorgegebenen und in der Betriebsanweisung Flugverkehrskontrolle (BA-FVK; DFS, 2006) übernommenen Zielsetzung der „Vermeidung von Zusammenstößen am Boden". Die Erfüllung dieser Pflicht ist wiederum durch konkrete Handlungsausführungen sowie Denk- und Planungsprozesse nachzukommen. Ob im Ernstfall allerdings die Hauptverantwortlichkeit für die Verhinderung

[17] MANTEA: Management of Surface Traffic in European Airports
[18] ATHOS: Airport Tower Harmonised Controller System
[19] VICTOR: Virtual Control Tower Research Studies
[20] Für eine genaue Definition von Runway Incursions sei auf Huber (2012) verwiesen.

von Runway Incursions beim Lotsen oder beim Piloten liegen sollte, ist eine schwierige Frage und wird unter anderem bei Huber (2012) diskutiert. Für die vorliegende Arbeit wird davon ausgegangen, dass diese Pflicht insoweit gilt, als dass durch die Ausführung der nötigen Aufgaben alle vorbeugenden Maßnahmen ergriffen werden sollten, um eine gefährliche Annäherung zweier Flugzeuge auf dem Rollfeld zu verhindern. Dies gilt gleichermaßen für Flugzeuge und Bodenfahrzeuge. Es können grundlegend zwei Arbeitsplätze bzw. Positionen unterschieden werden: TWR-Position (Platzverkehrslotse), welche für die Abwicklung von für Starts und Landungen zuständig ist, sowie die GND-Position (Rollverkehrslotse) welche für den Rollverkehr zuständig ist (vgl. Kapitel 4.4).

4.3.2 Beschreibung der Aufgabenanalysen

Für die Generierung eines umfassenden Bildes und eines Status-Quo der Aufgaben von Towerlotsen wurden die folgenden Aufgabenanalysen in die engere Auswahl einbezogen. Da bei einigen Analysen die Originaldokumente nicht mehr erhältlich sind, ist jeweils angegeben, wo auf Sekundärliteratur zurückgegriffen werden musste. Es wurden die folgenden Aufgabenanalysen in eine engere Auswahl einbezogen, da sie sich speziell mit dem Towerlotsen beschäftigen:

- Analyse aus dem EUROCONTOL-Projekt MANTEA (Rossi et al., 1996); (Sekundärliteratur: Buck et al., 1996; Marti, 1998, 2000; Tavanti, 2006a)
- Analyse und dem EUROCONTROL-Projekt ATHOS (Courboulay & Kahn, 1996); (Sekundärliteratur: Buck et al., 1996; Marti, 1998, 2000; Tavanti, 2006a)
- Analyse der FAA; (Primärliteratur: Alexander et al., 1989; Ammerman, Fairhurst, Hostetler, & Jones, 1988; Ammerman, Becker, et al., 1988)
- Analyse aus dem DFS-Projekt VICTOR; (Primärliteratur: Human-Factors-Consult, 2009)
- Analyse des Royal Air Force Institute of Aviation Medicine; (Primärliteratur: Cox, 1994a, 1994b)

Im Folgenden werden die einzelnen Analysen genauer beschrieben, sowie deren Verwendbarkeit analysiert.

4.3.3 Analyse aus dem EUROCONTROL-Projekt MANTEA

Die Analyse des MANTEA-Projektes ist in Tavanti (2006a) beschrieben. Das Ziel dieses Projektes war die Entwicklung von Entscheidungsunterstützungssystemen zur Verbesserung des Managements des Oberflächenverkehrs an Flughäfen. Das MANTEA-Projekt beinhaltet außerdem die Automatisierung von Funktionen wie der Planung des Oberflächenverkehrs und Überwachung der Übereinstimmung mit Instruktionen, die Erkennung potentieller Konflikte und ihre Lösung (Tavanti, 2006a). Die MANTEA-Studie analysiert die beiden Positionen TWR

und GND. Die Studien wurden in den zwei Towern Rom-Fiumicino und Amsterdam-Schipol durchgeführt. Hierbei wurde nach der typischen Zerlegung von Hauptaufgaben in Unteraufgaben vorgegangen, wie es bei einer hierarchischen Aufgabenanalyse üblich ist (Tavanti, 2006a). Für die Erstellung eines generalized task models (GTM) wurde die Methode MUSE (Method for Usability Engineering) verwendet (Lim & Long, 1994).

Die Dekompositionsgrafen beider Positionen weisen einen ähnlichen, aber recht geringen Detailgrad auf. Es sind daher keine konkreten Rückschlüsse auf detaillierte Handlungen oder kognitive Prozesse möglich. Im Gegensatz zu Cox (1994a, 1994b, siehe unten) liefert die Analyse des MANTEA-Projekts eher einen groben Überblick über die allgemeinen Tätigkeiten des Lotsen, als eine detaillierte Beschreibung der Arbeitsabläufe.

4.3.4 Analyse aus dem EUROCONTROL-Projekt ATHOS

Das ATHOS-Projekt lief von 1996 bis 2000 und hatte die Ausarbeitung einer neuen Mensch-Maschine-Schnittstelle für Flughafentower zum Ziel, die in zukünftige A-SMGCS-Systeme[21] integriert werden sollte. Die Daten der ATHOS-Studien wurden an fünf repräsentativen Flughäfen erhoben (Buck et al., 1996; Courboulay & Kahn, 1996; Marti, 1998, 2000): Paris-Charles-De Gaulle, Madrid-Barajas, Palma de Mallorca, Amsterdam Schipol und Frankfurt a. M.. Nach einer Analyse vor Ort wurde auch hier ein Task Analysis Model erstellt. Allerdings weist bereits Tavanti (2006a) darauf hin, dass über die verwendete Methodologie zur Aufgabenanalyse nur eine sehr kurze Stellungnahme existiert. Sie geht aber davon aus, dass dieselbe Methodik verwendet wurde, wie für das MANTEA-Projekt. Weiterhin war nur der Report auffindbar, der die Aufgabenanalyse des Flughafens Paris-Charles-De-Gaulle beschreibt. Ein weiterer Mangel des Dokumentes ist das Fehlen einer ausführlichen Beschreibung der Aufgaben, sodass diese oft nicht völlig verständlich sind. Ebenso ist für das ATHOS-Projekt nur die Position des TWR beschrieben. Auch diese Analyse weist einen ähnlichen Detailgrad wie die des MANTEA-Projektes auf.

4.3.5 Analyse der FAA (FAA Air Traffic Control Operations Concepts)

Die Analyse der FAA umfasst sieben Bände, wobei der fünfte gänzlich den Towerlotsen gewidmet ist und sich auf Controller in der sogenannten Tower Control Computer Complex (TCCC)-Umgebung bezieht (Ammerman, Becker, et al., 1988). Band sieben beschäftigt sich mit Lotsen in der damaligen aktuellen Tower-Umgebung (Alexander et al., 1989). Im ersten Band werden der Zweck der Analyse, sowie der betriebliche Hintergrund und die Methodik, nach der die Studien durchgeführt wurden, beschrieben (Ammerman, Fairhurst, et al., 1988). Alle Analysen wurden im amerikanischen Flugverkehrskontrollsystem durchgeführt und beziehen

[21] Advanced Surface Movement Guidance and Control System – Ein Kontrollsystem, welches den Lotsen beim Kontrollieren und Steuern des Verkehrs am Boden eines Flughafens unterstützen soll.

sich somit auch auf Lotsen im amerikanischen Raum. Eine Übertragbarkeit der identifizierten Aufgaben, der verwendeten Hilfsmittel, sowie der Abläufe und Aufteilungen der Verantwortungen in einzelnen Positionen auf den europäischen Luftraum ist damit nicht anzunehmen, da sich die Regulationen teilweise stark unterscheiden können. Weiterhin wurde diese Studie bereits 1988 durchgeführt, und es kann davon ausgegangen werden, dass das Flugverkehrskontrollsystem seither starken Veränderungen unterworfen war, sei es durch technische Neuerungen, durch Zunahme des zu kontrollierenden Luftverkehrs oder auch durch Änderungen der Aufgaben und Rollenverteilungen. Aus diesen starken Einschränkungen ergab sich die Notwendigkeit, die Analyse aus der weiteren Betrachtung auszuschließen.

4.3.6 Analyse aus dem DFS-Projekt VICTOR

Die Studie von Human-Factors-Consult (2009) besitzt, wenngleich sie sehr aktuell ist, ebenfalls einige Einschränkungen: Sie basiert nach eigenen Angaben auf den Ergebnissen des MANTEA und ATHOS-Projektes, welche aber mit eigenen Beobachtungen und Datenerhebungen an den Regionalflughäfen Niederrhein und Dortmund verglichen wurden. Da einige Unterschiede in der Aufgabenverteilung an Regionalflughäfen und größeren Flughäfen bestehen: *„ist zu beachten, dass sich die identifizierten Aufgaben bei dieser Aufgabenanalyse auf die Arbeitsrealität der Fluglotsen in Niederrhein und Dortmund bezogen, wo teilweise keine klare Aufgabentrennung von Platz- und Rollverkehrslotse bestand."* (Human-Factors-Consult, 2009, S. 5). Es wurde daher auch keine Trennung zwischen Rollverkehrs- und Towerlotse vorgenommen. Zusätzlich wird darauf hingewiesen, dass das Ziel darin bestand *„…ausschließlich die Aufgaben systematisch zu analysieren, bei denen visuelle Informationen aus der Umwelt bezogen werden."* (Human-Factors-Consult, 2009, S. 5). Somit fokussiert die Aufgabenanalyse weniger darauf, welche Handlungen konkret vollzogen werden müssen (durch welche konkreten Handlungsschritte oder Denkprozesse), sondern mehr darauf, an welchen Objekten sie zu vollziehen sind. Die Aufgabenanalyse fließt in den Modellerstellungsprozess daher insofern ein, als dass sie einerseits ein allgemeines Bild der Arbeit von Towerlotsen vermitteln kann und andererseits Hinweise darauf enthält, aus welchen Informationsquellen der Lotse welche Informationen zieht und welche Hilfsmittel er verwendet, um seine Aufgaben auszuführen.

Es können zwei Hauptaufgaben als wichtig für die Modellierung herausgestellt werden:

(1) Beobachtung aller sichtbaren (Flug-) Bewegungen auf den Betriebsflächen und in der Umgebung des Flughafens;
(2) Erteilung notwendiger Freigaben und Anweisungen an Luftfahrzeuge.

Für die Hauptaufgabe 1 konnte festgestellt werden, dass Fluglotsen für die Ausführung ihrer Aufgaben immer die gleichen Basisinformationen benötigen: Typ/Identifizierung (Flugzeugtyp; Rufzeichen), Position, Lage/Orientierung (z.B. Ausrichtung der Nase), Zustand (z.B. Triebwerke,

Fahrwerk) und Verhalten (Bewegungsrichtung, Beschleunigung). Für die Hauptaufgabe 2 wurde festgestellt, dass die Kontrolle der Objekte hauptsächlich auf deren aktuellem Aufenthaltsort beruht und nicht so sehr vom Bewegungszustand abhängt.

4.3.7 Analyse des Royal Air Force Institute of Aviation Medicine

Das Hauptziel der Analyse von Cox (1994a, 1994b) war die Unterstützung von Risikoanalysen in der Flugverkehrskontrolle. Sie sollte die nötigen Informationen liefern, um Fehlerentstehungen besser einschätzen zu können. Sowohl die eigentlichen Analysen und die Dekomposition der Aufgaben als auch der Report selbst sind sehr ausführlich und detailliert. Die Analyse wurde in drei Teile unterteilt: den Hauptreport (Volume 1: Main report; Cox, 1994a), die Anhänge mit den genauen Dekompositionsgraphen (Volume 2: Appendices; Cox, 1994b), sowie die eigentliche Zuverlässigkeitsanalyse (Volume 3), die höchstwahrscheinlich nie veröffentlicht wurde.

Im Hauptreport wird genau auf die Analysemethode und die Art und Weise der Datenerhebung eingegangen, sowie die Aufgaben sehr ausführlich beschrieben. Es werden die vier Positionen en-rounte, TMA[22], sowie Approach und Departure (TWR; wobei nochmals zwischen dual und single runway operation unterschieden wird[23]), sowie Ground Movement Controller (GND) behandelt. Zusätzlich wird noch auf die Lösung von Konflikten und die Reaktionen bei Notfällen eingegangen. Der Autor hielt sich hierbei so genau wie möglich an die Methodik der hierarchischen Aufgabenanalyse. Die Informationen wurden durch Cox selbst durch relativ kurze Besuche in den verschiedenen von der NATS[24] betriebenen Einrichtungen gesammelt. Die Analysen der Towerlotsen-Aufgaben wurden sowohl in London Heathrow (dual runway operation) als auch London Gatwick (single runway operation) durchgeführt. Die Beobachtungen erfolgten außerdem zu verschiedenen Tages- und Nachtzeiten. Zusätzlich wurden semi-strukturierte Interviews durchgeführt, um einzelne Aufgabenelemente zu verdeutlichen, die beobachtet wurden und um Informationen zu Situationen zu erhalten, die nicht beobachtet wurden. Zusätzliche Informationen zog Cox aus Journal-Artikeln und der bereits genannten Analyse der FAA (Alexander et al., 1989; Ammerman, Becker, et al., 1988). Letztere verwendete er allerdings nur als allgemeine Richtlinie.

Wie beschrieben wurde die TWR-Position sowohl für dual als auch single runway operations untersucht. Da in der vorliegenden Arbeit davon ausgegangen wird, dass ein Lotse die Kontrolle des Flughafens vornimmt, werden die bei Cox für single runway operation beschriebenen

[22] TMA: Terminal Manoeuvring Area; auch Terminal Control Area (TCA) genannt: Kontrolle des Luftverkehrs in einiger Umgebung um den Flughafen, in dem Sink- und Steigflüge vorgenommen werden.
[23] Dual-Runway operation besteht an Flughäfen, mit zwei parallelen Pisten, wobei eine Piste nur für Starts, die andere nur für Landungen verwendet wird. Single-Runway operation besteht an Flughäfen mit nur einer Piste, wobei diese sowohl für Starts als auch Landungen verwendet wird.
[24] National Air Traffic Services; Flugsicherungsorganisation Großbritanniens

Aufgaben verwendet. Für die Position des *Towerlotsen (TWR)* werden sieben Hauptaufgaben beschrieben (Cox, 1994b, Anhang J, S.1):

(1) Formulierung eines Plans zur Integration der Bewegungen der Ankünfte und Abflüge (*„Forlulate plan to integrate arrivals and departures"*)
(2) Führung der Abflüge (*„Conduct aircraft departures"*)
(3) Führung der Ankünfte (*„Conduct aircraft arrivals"*)
(4) Durchführung von Routineaufgaben *(„perform routine 'maintenance' tasks"*)
(5) Übergabe der Kontrolle über die Abflüge an die nächste Behörde (*„Transfer control of departures to next agency"*)
(6) Übergabe der Kontrolle über die Ankünfte an den Rollverkehrslotsen (*„Transfer control of arrivals to Ground Movement Controller"*)
(7) Handlungen nachdem Flugzeug Zuständigkeitsbereich verlassen hat (*„Perform actions after aircraft has left tower control")*

Grundsätzlich muss dieser Lotse die Landungen und Starts sicher und effizient integrieren, um eine maximale Nutzung der Start-/Landebahn zu erreichen. Dazu kann er verschiedene sogenannte Conditional Clerances (bedingte Freigaben) vergeben (z.B. „…after next landing, line up and wait Runway 25R"). Weiterhin ist an Flughäfen eine zweite Lotsenposition besetzt, die für den Bodenverkehr zuständig ist (GND). Für die *GND-Position* beschreibt Cox acht Hauptaufgaben (Cox, 1994b, Anhang O, S.1):

(1) Annahme von Flugzeugen von der vorhergehenden Position (*„Accept aircraft from previous control position"*)
(2) Formulierung eines Plans zur Integration der Bewegungen der Ankünfte und Abflüge (*„Forlulate plan to integrate movements of arrivals and departures"*)
(3) Führung der Ankünfte zum Stand (*„Conduct direction of arrivals to stand")*
(4) Führung der Abflüge zur Startbahn (*„Conduct direction of departures to (departures) runway"*)
(5) Führung gezogener (Luft-)Fahrzeuge (*„Conduct direction of towed aircraft"*)
(6) Durchführung von Routineaufgaben (*„Perform routine 'maintenance' tasks"*)
(7) Bearbeitung von Notfällen am Boden (*„Conduct handling of ground emergency"*)
(8) Übergabe der Kontrolle der Abflüge an die nächste Behörde (*„Transfer control of departures to next agency"*)

Die hier beschriebenen Hauptaufgaben stellen die obere Ebene der jeweiligen Dekompositionsgrafen dar und sind in zahlreiche Unteraufgaben und kleinere Handlungsschritte unterteilt (vgl. Kapitel 4.5). Die Analyse von Cox (1994a, 1994b) weist einen hohen Detailgrad auf, weshalb hauptsächlich diese, verwendet wurde, um das theoretische

Modell (Kapitel 4.5) und somit einen prototypischen Arbeitsablauf von Towerlotsen als integratives Prozessmodell zu erstellen (vgl. Smieszek, Huber, & Jürgensohn, 2011).

4.4 Feldstudie der Arbeitsumgebung

Für die Studie der Arbeitsumgebung wurde je ein Besuch im Flughafentower Berlin Schönefeld im Mai 2011, sowie im Flughafentower Braunschweig-Wolfsburg im Februar 2013 durchgeführt, um einen Eindruck der Arbeit im Tower zu erhalten, und die Möglichkeit zu haben, den Lotsen persönlich Fragen zu stellen. Es wird zunächst allgemein auf die Abläufe und die Zuständigkeiten der Towerlotsen eingegangen. Anschließend wird kurz auf die zur Verfügung stehenden Informationsquellen und Hilfsmittel eingegangen, bevor die einzelnen Positionen TWR und GND, sowie die allgemeinen Abwicklungsprioritäten beschrieben werden.

Zusätzlich wurde eine Analyse zahlreicher Literatur zur Flugverkehrskontrolle durchgeführt (DFS, 2006; Durso & Manning, 2008; Federal Aviation Administration, 2011; ICAO4444, 2007; Mensen, 2004; Merz, 2007; Nolan, 2011; Pinska & Bourgois, 2005; Pinska, 2008; Stubbe, 2011; Tavanti, 2006b). Hierbei wurden die Zuständigkeiten der Lotsen und die damit verbundenen Pflichten, die zur Aufgabenerfüllung zur Verfügung stehenden Hilfsmittel, sowie die allgemeinen Regularien und Beschränkungen der Flugverkehrskontrolle und die benötigten Informationen analysiert.

4.4.1 Allgemeine Abläufe und Zuständigkeiten

Abbildung 7 stellt die Hauptphasen des finalen Anfluges und die zugehörigen Tätigkeiten der Towerlotsen grafisch dar. Flugzeuge werden von der Ankunft/Abflug-Position (APP; Final Approach/ Departure unit) an den TWR übergeben, welcher für Starts und Landungen zuständig ist. Anschließend werden sie an den GND, der für den Bodenverkehr zuständig ist, weitergegeben. Der Zuständigkeitsbereich der Towerlotsen umfasst den gesamten Flughafen. Eine genaue Trennung der Zuständigkeiten zwischen den Tower-Positionen und der vorgeschalteten Approach-Position (zuständig für die Koordination der An- und Abflüge in weiterer Umgebung um den Flughafen) ist nicht streng definiert und hängt von lokalen Gegebenheiten ab (ICAO-9426, 1984; Tavanti, 2006b). Am Stockholmer Flughafen Arlanda beispielsweise hat der Bereich, für den der Towercontroller zuständig ist, eine Größe von 20NM (Nautische Meilen) um den Flughafen herum (Tavanti, 2006b). In den meisten Fällen gibt es sogenannte Letters of Agreement zwischen den Approach- und den Tower-Positionen, in denen die Zuständigkeitsbereiche beschrieben werden (Nolan, 2011). Die Hauptaufgabe von Flughafenverkehrslotsen sind in der ICAO4444 (2007) definiert:

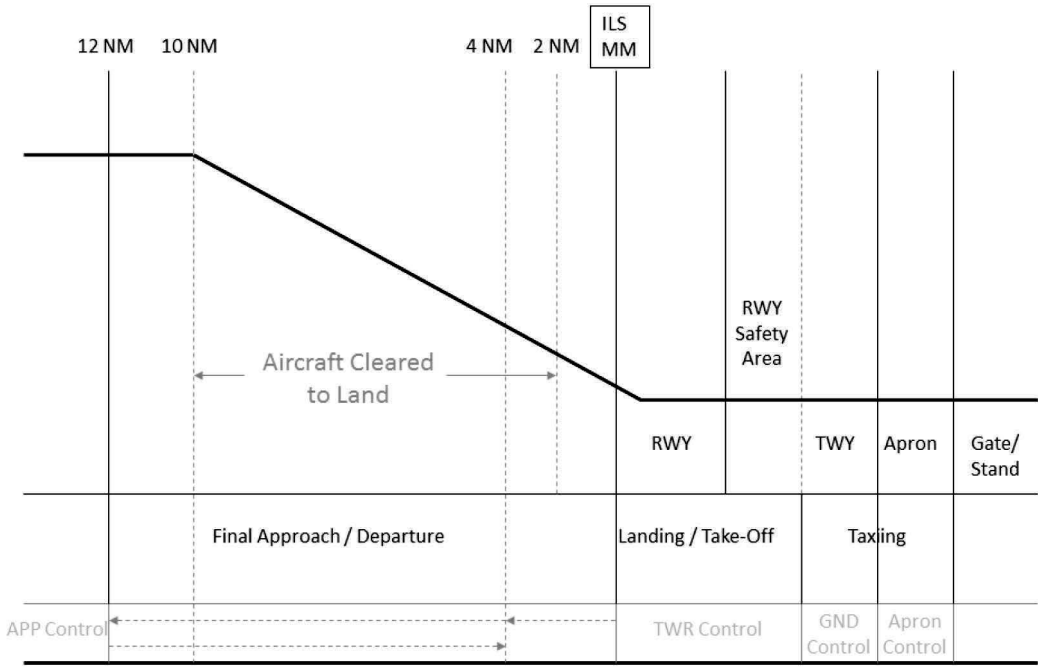

Abbildung 7: Phasen des Finalen Anfluges und die dazugehörigen Aufgaben des Towerlotsen (eigene Darstellung nach Tavanti, 2006b)

„Aerodrome control towers shall issue information and clearances to aircraft under their control to achieve a safe, orderly and expeditious flow of the air traffic on and in the vicinity of an aerodrome with the object of preventing collision(s) between:

a. *aircraft flying within the designated area of responsibility of the control tower, including the aerodrome traffic circuits;*
b. aircraft operating on the manoeuvring area;
c. aircraft landing and taking off;
d. aircraft and vehicles operating on the manoeuvring area;
e. aircraft on the manoeuvring area and obstructions on that area" (S. 7-1)

Dabei ist das sogenannte Manoeuvring Area *„that part of an aerodrome to be used for take-off, landing and taxiing of aircraft, excluding aprons"* (ICAO4444, 2007, S.1-11). Damit ist die Manoeuvring Area von der Movement Area zu unterscheiden, welche zusätzlich die Aprons[25] beinhaltet. Diese Unterscheidung ist wichtig im Hinblick auf die Zuständigkeiten der Lotsen, denn nur die Manoeuvring Area ist abhängig von der Aufsicht und Führung des Lotsen. Nach Bergé (2005) sollte die Aufsicht allerdings auch Flugzeuge auf dem Vorfeld mit einbeziehen, da der Lotse die Position des Flugzeuges kennen muss, um spätere Konflikte auf der Manoeuvring Area zu verhindern. Auch in der ICAO4444 wird eine solche Überwachung gefordert: *„Aerodrome controllers shall maintain a continuous watch on all flight operations on and in the vicinity of an aerodrome as well as vehicles and personnel on the maneuvering area. Watch shall be*

[25] Apron: Vorfeld; Abfertigungs-, Rangier-, Abstell- und Wartungsfläche für Luftfahrzeuge. Diese können auch weiter entfernt vom Terminal liegen.

maintained by visual observation, augmented in low visibility conditions by an ATS surveillance system when available" (ICAO4444, 2007, S.7-1).

Für die Modellierung bedeutet dies, dass prinzipiell zwei Positionen modelliert werden müssen. Aufgrund der bereits in Kapitel 4.2 beschriebenen Einschränkung soll aber nur ein Lotse (bzw. ein Lotsenmodell) die Aufgaben beider Positionen übernehmen.

4.4.2 Hilfsmittel zur Informationsaufnahme und Aufgabenbearbeitung

Abbildung 8: Der Kontrolltower Berlin Tegel von innen (Bildquelle: DFS; http://www.dfs.de/dfs_homepage/mediathek/de/Defaults/Mediatheken/DFS-Fotos%3A/04%2006%2011%20KE%200519-01_18x24.jpg)

Zur Aufgabenausführung steht den Lotsen eine Reihe von Hilfsmitteln zur Verfügung. Diese sollen im Interaktionsmodell modelliert werden. Eines der wichtigsten Werkzeuge der Fluglotsen allgemein sind die sogenannten *Kontrollstreifen oder Flightstrips* (vgl. Abbildung 9). Kontrollstreifen *„constitute a highly flexible means of planning departure sequences and can be written on, re-ordered and otherwise physically manipulated in this process"* (Fields, Amaldi, & Tassi, 2003 nach Tavanti, 2006b). Diese Kontrollstreifen stellen eine Repräsentation jedes einzelnen Flugzeuges dar. Die Strips enthalten zahlreiche „statische" Informationen über den Flug, wie beispielsweise das Rufzeichen, die geschätzte Ankunfts- oder Abflugszeit, woher das Flugzeug kommt, wohin es fliegt, wann es abfliegt uvm. Zusätzlich können per Hand wichtige Informationen des Verhaltens der Flugzeuge hinzugefügt werden (z.B. Freigaben, Position am Boden; Marti, 2000). Jeder Kontrollstreifen wird dem betreffenden Lotsen einige Minuten, bevor das Flugzeug Kontakt aufnimmt und der Lotse die Kontrolle übernimmt, zugestellt. Es gibt jeweils einen Flugstreifen pro Flugzeug und das physische Weiterreichen von einer Position zur

nächsten kennzeichnet den Transfer der Verantwortlichkeit für das Flugzeug. Die Kontrollstreifen werden in der für den jeweiligen Lotsen angemessenen Reihenfolge auf einem sogenannten Stripboard (oder Stripbay) platziert. Somit beschreibt die Reihenfolge der Strips auf dem Board die Reihenfolge der Abarbeitung des Verkehrs (Marti, 2000). *„The strip board is interesting as it provides a picture of the situation at hand, a plan of the traffic evolution and the associated estimate of the controlled workload, and a memory of activity"* (Marti, 2000, S.68). Die Flugstreifen dienen damit der Externalisierung von Informationen. Das bedeutet, dass nicht alle Informationen zu Flugzeugen im Arbeitsgedächtnis aufrechterhalten werden müssen, da sie auf dem Flugstreifen enthalten sind und jederzeit von dort abgerufen werden können. In einigen Kontrolltowern sind mittlerweile elektronische Flugstreifen im Einsatz (u.a. in Berlin Schönefeld), die die Übergabe von einer Position zur anderen immens vereinfachen, da sie per EDV übermittelt werden können und nicht mehr händisch übergeben werden müssen. Für weitere Ausführungen zu den möglichen Auswirkungen der Umstellung von Papier auf elektronische Flugstreifen sei auf Berndtsson & Normark (1999), Durso, Batsakes, Crutchfield, Braden, & Manning (2004), Durso & Manning (2008), Mackay (2000), sowie Truitt (2005, 2006) verwiesen.

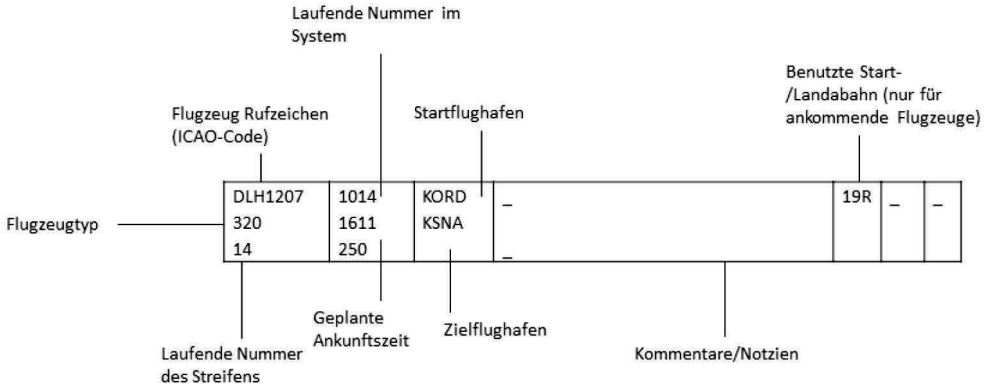

Abbildung 9: Beispiel für einen Flugstreifen (eigene Darstellung)

Ein Beispiel für einen einfachen Papierstreifen, wie er auch im Validierungsversuch verwendet worden ist, ist in Abbildung 9 dargestellt. Jeder Teil des Flugstreifens trägt eine spezifische Information über das Flugzeug. Außerdem ist der Streifen *„[…] a shared note pad conveying to the members of the team what actions have been taken with respect to particular aircraft […]"* (Bentley et al., 1992, S. 126).

Als zweites wichtiges Hilfsmittel und primäre Informationsquelle dient *die Außensicht*, über die alle Informationen über die aktuelle Verkehrssituation durch einen Blick auf den Flughafen erhalten werden können. Zur direkten *Kommunikation* mit den Führern der betreffenden Luftfahrzeuge steht außerdem ein *Funkgerät* zur Verfügung. Weitere Hilfsmittel sind das *Bodenradar*, welches die Flugzeugbewegungen am Boden darstellt, ein *Luftradar* auf dem landende und alle Flugzeuge in der näheren Umgebung des Flughafens dargestellt werden (für

den TWR controller), ein *Flight Data Processing System (FDP)*, welches diejenigen Daten anzeigt, die den jeweiligen Flug betreffen (z.B. Rufzeichen, Flugzeugtyp, time slot, zugeordnete RWY usw.), ein *Interface zur Steuerung der Beleuchtungssysteme*, ein *Wetterinformationsdisplay*, sowie ein *Telefon*. (Tavanti, 2006b; vgl. auch Manske, von Schlippe, & Möhlenbrink, 2013). Für die Modellierung werden hierbei nur die primären Informationsquellen Außensicht, Funkkommunikation und Flugstreifen betrachtet, da diese als am Wichtigsten anzusehen sind. Alle anderen Hilfsmittel sind eher von geringerer Wichtigkeit (vgl. Pinska & Bourgois, 2005; Pinska & Tijus, 2007; Pinska, 2006, 2008).

4.4.3 POSITION TOWERLOTSE (TWR)

Neben den Hilfsmitteln sind die Zuständigkeiten der beiden bereits beschriebenen Positionen für die Modellierung von Bedeutung, um zu definieren, auf welche Bereiche sich die Verantwortung erstreckt und welche Besonderheiten es zu beachten gilt. Die Bezeichnung der Position des Towerlotsen (TWR) ist dabei etwas irreführend, denn alle Lotsen am Flughafen sitzen schließlich im Tower. Ist allerdings von der Position TWR die Rede, sind diejenigen Lotsen gemeint, die speziell für die Handhabung von Ankünften und Abflügen, sprich Starts- und Landungen zuständig sind. Laut der ICAO4444 (2007) ist die TWR-Position *„normally responsible for the operations on the runway and aircraft flying within the area of responsibility of the aerodrome control tower"* (S.7-1). Das bedeutet, dass die Aufgabe der TWR-Position darin besteht, ankommende Flugzeuge in einen flüssigen und geordneten Ablauf zu bringen und ausgehenden Verkehr in diesen Ablauf einzuordnen, sodass ausreichend Staffelung zwischen den landenden und startenden Flugzeugen eingehalten wird (Nolan, 2011). Der TWR-Lotse ist zuständig für: 1) die landenden Flugzeuge bis diese die Landebahn verlassen, 2) die ausgehenden Flüge von der Halteposition zum Start bis das Flugzeug zur APP-Position übergeben wird (Tavanti, 2006b). Der TWR ist die erste Anlaufstelle für ankommende Flüge direkt nach der APP-Position. Er gibt die Landefreigabe („cleared to land") bzw. die Startfreigabe („cleared for take off") bei ausgehendem Verkehr, sowie Informationen über Wetter- und Windverhältnisse. Der TWR muss die Zustände und den Status aller Start-/Landebahnen kennen. Der Belegungsstatus der Piste wird durch direkte Beobachtung (Außensicht) und das Radar kontrolliert. Hat ein ausgehendes Flugzeug den Zuständigkeitsbereich des Flughafens verlassen, leitet der TWR das Flugzeug zur APP-Position weiter. In Koordination mit der APP-Position ist der TWR für die Definition der angemessenen Runway-Konfiguration verantwortlich: *„the runway or runways that, at a particular time, are considered by the aerodrome control tower to be the most suitable for use by the types of aircraft expected to land or take off at the aerodrome"* (ICAO4444, 2007, S.7-2). Außerdem ist diese Position für die Erstellung der Sequenz ausgehender Flüge zuständig (für die Erstellung der Sequenz ankommender Flüge ist die APP-Position zuständig.). Für diese Aufgabe ist die Kenntnis der

sogenannten Wirbelschleppenkategorie des Flugzeugs oder dessen Typ, sowie der Beziehungen zu den vorausgegangenen und nachfolgenden Flugzeugen von enormer Wichtigkeit. Nach der ICAO4444 (2007, S.4-11) werden drei Kategorien von Flugzeugen unterschieden:

(1) HEAVY (H) – alle Flugzeuge mit einem Gewicht von 136.000 kg und mehr
(2) MEDIUM (M) – alle Flugzeuge mit einem Gewicht von weniger als 136.000 kg, aber mehr als 7.000 kg
(3) LIGHT (L) – alle Flugzeuge mit einem Gewicht von 7.000 kg oder weniger.

Die Staffelung zwischen Flugzeugen berücksichtigt das Auftreten dieser Wirbelschleppen. Daher werden abhängig vom Flugzeugtyp, der Runway-Konfiguration und der Route, die das Flugzeug fliegen wird, unterschiedliche Abstände zwischen den Flugzeugen eingehalten. In Abhängigkeit der genannten Einschränkungen bestehen zahlreiche Regeln, welche die Wirbelschleppenstaffelung bestimmen. Eine ausführliche Beschreibung der Wirbelschleppenstaffelungsminima können in ICAO4444 (2007) nachgelesen werden. Offene Interviews von Tavanti (2006b) mit Lotsen des Fiumicino Flughafens in Rom weisen außerdem darauf hin, dass der TWR-Lotse auch darauf achtet, dass ausreichend Abstand zwischen den ankommenden Flugzeugen eingehalten wird.

4.4.4 Abwicklungsprioritäten und Staffelungsminima des TWR

Die Priorität der Bearbeitung von Ankünften und Abflügen liegt eindeutig auf ankommendem Verkehr: *„An aircraft landing, or in the final stages of an approach to land shall normally have priority over an aircraft intending to depart from the same or an intersecting rumway“* (ICAO4444, 2007, S.7-13). Ferner dürfen Flugzeuge nicht näher an eine Start-/Landebahn heranrollen als bis zum Haltepunkt. Für Luftfahrzeuge, die dieselbe Piste benutzen, gelten folgende Regeln (DFS, 2002):

„Startende Luftfahrzeuge sind von anderen, dieselbe Piste benutzenden Luftfahrzeugen zu staffeln, indem sichergestellt wird, dass sie ihren Startlauf nicht beginnen, bevor eine der folgenden Bedingungen erfüllt ist: Das vorher gestartete Luftfahrzeug hat das Ende der Piste überflogen oder hat abgehoben und eine Kurve eingeleitet, die eine Staffelungsunterschreitung ausschließt; das vorher gelandete Luftfahrzeug hat die Piste verlassen.“ (S.32-2).

„*Landende Luftfahrzeuge sind von anderen, dieselbe Piste benutzenden Luftfahrzeugen zu staffeln, indem sichergestellt wird, dass sie den Anfang der Piste nicht überfliegen, bevor eine der folgenden Bedingungen erfüllt ist: Das vorher gestartete Luftfahrzeug hat das Ende der Piste überflogen oder hat abgehoben und eine Kurve eingeleitet, die eine Staffelungsunterschreitung ausschließt; das vorher gelandete Luftfahrzeug hat die Piste verlassen.“* (S. 32-2)

Wenn ein Flugzeug landet, dürfen andere Flugzeuge solange nicht am Ende einer aktiven Runway aufrollen und halten, bis das landenden Flugzeug den Haltepunkt passiert hat (Tavanti,

2006b). Ein landendes Flugzeug darf solange das Ende der Runway nicht passieren, bis das vorhergehende abfliegende Flugzeug das Ende der Runway überschritten hat oder begonnen hat zu drehen oder das vorhergehende landende Flugzeug die Runway verlassen hat. Ebenso darf ein abfliegendes Flugzeug nicht mit dem Startlauf beginnen, bis das vorausgehende abfliegende Flugzeug das Ende der Runway passiert hat oder begonnen hat zu drehen oder das vorausgehende landende Flugzeug die Runway verlassen hat (Tavanti, 2006b).

Wenn ein Flugzeug gelandet ist und die Runway verlassen hat, übergibt der TWR das Flugzeug zur GND-Position. Die Übergabe ist gekennzeichnet durch die Kommunikation mit dem Piloten, der aufgefordert wird die Funk-Frequenz zu wechseln und GND zu kontaktieren, sowie durch den physischen Transfer des Kontrollstreifens von TWR zu GND. Derselbe Ablauf wird bei abfliegendem Verkehr mit der APP-Position eingehalten (Tavanti, 2006b).

4.4.5 Position Rollverkehrslotse (Ground; GND)

Diese Position ist nach ICAO4444 (2007, S. 7-1) *„normally responsible for traffic on the manoeuvring area with the exception of runways"*. Sie ist also für die Staffelung von Flugzeugen und Fahrzeugen, welche auf dem Vorfeld, den Rollwegen und inaktiven Start- und Landebahnen operieren, verantwortlich. Dies erstreckt sich auf Flugzeuge, die zum Start rollen, Flugzeuge, die nach der Landung zum Terminal rollen und alle Bodenfahrzeuge auf der Movement Area (Nolan, 2011). Der Pilot kontaktiert den Rollverkehrslotsen für die Push-Back- und die Rollfreigabe. Nach der Zustimmung des Rollverkehrslotsen wird das Flugzeug durch die Manoeuvring Area geleitet bis es den Haltepunkt nahe der Start-/Landebahn erreicht hat. Anschließend wird das Flugzeug an den Towerlotsen übergeben (Tavanti, 2006b). Der Rollverkehrslotse überwacht und leitet alle Bewegungen am Boden. Er übermittelt dem Piloten die Rollrouten und achtet dabei darauf, dass Kollisionen mit anderen Flugzeugen oder Objekten vermieden und das Risiko, dass Flugzeuge auf eine aktive Start-/Landebahn, aufrollen minimiert werden. Der Rollverkehrslotse weist sowohl ankommenden als auch abfliegenden Flügen Prioritäten zur Rollwegsnutzung zu. Die Position von Flugzeugen auf der Manoeuvring Area kann der Rollverkehrslotse durch direkte Beobachtung über die Außensicht, das Bodenradar oder durch Funkkommunikation bestimmen (er kann den Piloten nach dessen exakten Position fragen). Der Rollverkehrslotse kann einem Flugzeug erlauben auf eine aktive Start-/Landebahn aufzurollen, was allerdings mit dem Towerlotsen koordiniert werden muss und dessen Zustimmung bedarf (Tavanti, 2006b).

4.4.6 Prioritäten des GND

Grundlegend wird die Sequenz des GND nach der Regel „wer zuerst kommt mahlt zuerst" („first come, first served") festgelegt (Tavanti, 2006b). Oder mit anderen Worten: *„departures shall normally be cleared in the order in which they are ready to take off"* (ICAO4444, 2007, S. 7-

13). Ausnahmen von dieser Regel können gemacht werden, sofern damit die höchste Zahl an Abflügen mit der geringsten Durchschnittsverspätung erreicht wird, wobei allerdings bestimmte Faktoren berücksichtigt werden müssen (ICAO4444, 2007):

1) Flugzeugtypen und ihre relative Leistung
2) Routen, denen nach dem Start gefolgt werden muss
3) jedes festgelegte minimale Abflugsinterval zwischen den Starts
4) die Notwendigkeit, Wirbelschleppen-Staffelungsminima einzuhalten
5) Flugzeuge, denen Priorität zugesprochen werden sollte

Der Rollverkehrslotse kann in Übereinstimmung mit den Begrenzungen der Slots über die Abflugsequenz entscheiden und auf diese Weise Prioritäten zuweisen, um den Towerlotsen mit einer „eleganten" Sequenz zu versorgen.

4.5 Einordnung in und Erstellung des theoretischen Modells

Die Erstellung eines theoretischen Modells erfolgt durch Systematisierung der in Kapitel 4.3 vorgestellten Aufgabenanalyse anhand der Phasen der Handlungsregulation, die u.a. von Hacker (1986) formuliert wurden (vgl. Dörner & Schaub, 1995; Frese & Zapf, 1994)[26]. Zum besseren Verständnis der einzelnen Handlungen und Prozesse innerhalb der einzelnen Phasen, wurden die in den Aufgabenanalysen beschriebenen Haupt- und Unteraufgaben in einen prototypischen Arbeitsablauf der beiden Lotsenpositionen (TWR und GND) integriert. Hierbei wurde vor allem die Aufgabenanalyse von Cox (1994a, 1994b) verwendet, da diese den nötigen Detailgrad der hierarchischen Dekomposition aufweist (vgl. Kapitel 4.3.7). Die Handlungsregulationstheorie nach Hacker (1986) stellt die ausführlichste Theorie kognitiver Handlungsregulation dar und ist, anders als die meisten kognitiven Theorien und Informationsverarbeitungsansätze, durch ihre Verbundenheit mit dem Verhalten in realen Arbeitskontexten und Arbeitsresultaten eine starke Theorie zur makrokognitiven Modellierung (Frese & Zapf, 1994). Grundsätzlich kann der Handlungsprozess in sechs Phasen eingeteilt werden (vgl. Abbildung 10):

(1) *Zielbildung*

(2) *Informationssammlung und Orientierung*

(3) *Plangenerierung*

(4) *Entscheidung für eine Handlungsalternative*

(5) *Handlungsausführung*

[26] Eine ausführliche Beschreibung von Theorien der Entstehung menschlicher Handlungen ist nicht im Sinne dieser Arbeit. Daher sei hier auf Hacker (2005), Miller, Galanter und Pribram (1960), Rasmussen (1983), sowie für eine Übersicht auf Jürgensohn (1997), sowie Nerdinger, Blickler und Schaper (2008) verwiesen.

(6) *Feedbackverarbeitung*.

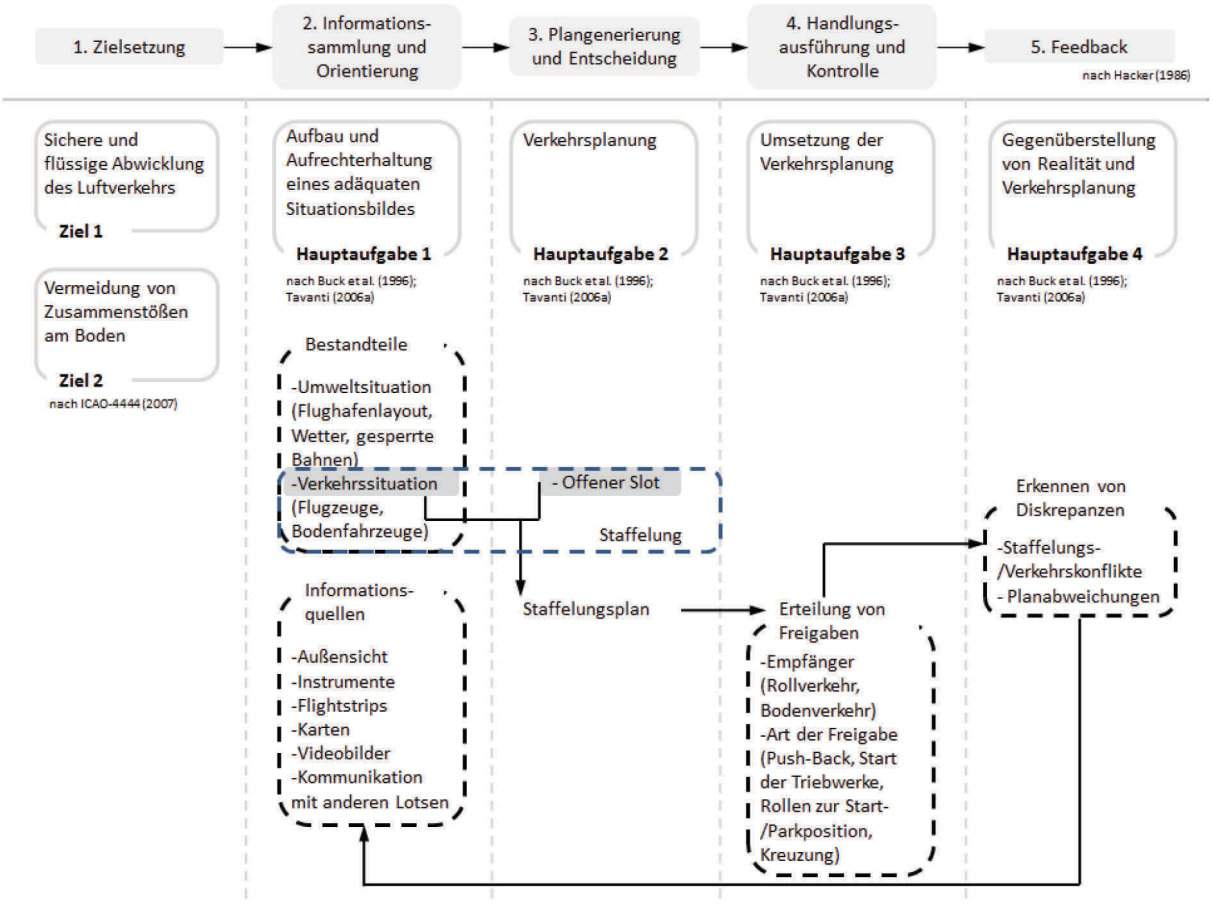

Abbildung 10: Aufgaben von Towerlotsen eingebettet in die Phasen der Handlungsregulation nach Hacker (1986) (nach Smieszek, Huber & Jürgensohn, 2011)

Die *Phase der Zielbildung* wird nicht vom Lotsen selbst durchgeführt. So ist das oberste Ziel in der Flugverkehrskontrolle mit der sicheren und flüssigen Abwicklung des Luftverkehrs und der Vermeidung von Zusammenstößen am Boden und in der Luft bereits vordefiniert (ICAO4444, 2007). Die Unterziele für den Lotsen werden durch die Anfragen von Flugzeugen auf die Erteilung bestimmter Freigaben definiert (Werther, Möhlenbrink, & Rudolph, 2007; Werther, 2006a). Die *Phase der Informationsaufnahme und Orientierung* beschreibt den Aufbau eines adäquaten Situationsbildes durch die Aufnahme der notwendigen Informationen aus der Umgebung. Dafür stehen dem Lotsen verschiedene Informationsquellen zur Verfügung, die bereits in Kapitel 4.4 beschrieben wurden. Hierdurch können Informationen über die aktuelle Verkehrssituation erhalten werden (vgl. Smieszek, Huber, & Jürgensohn, 2011; Smieszek & Karl, 2012; Smieszek, 2011). Auf Basis des generierten Situationsbildes erfolgt die Vorhersage und Planung der Verkehrssituation, welche in der *Phase der Plangenerierung* zur Erstellung eines Staffelungsplanes führt (Smieszek et al., 2011). In *der Phase der Entscheidungsfindung* erfolgt die Entscheidung für eine Handlungsalternative, welche anschließend in der *Phase der Handlungsausführung* durch die nötigen Aktionen implementiert wird. Die so entstehende neue Verkehrssituation wird anschließend in der *Phase der Feedbackverarbeitung* mit dem erstellten

Plan verglichen und eventuelle Diskrepanzen, wie Konflikte oder Planabweichungen, erkannt. Diese Phase erfolgt hierbei implizit wiederum durch die Aufnahme neuer Informationen und der Integration dieser in den neuen Handlungsablauf.

Im nächsten Schritt wurden die aus den Aufgabenanalysen extrahierten Haupt- und Unteraufgaben von Towerlotsen in die beschriebenen Handlungsphasen eingeordnet.

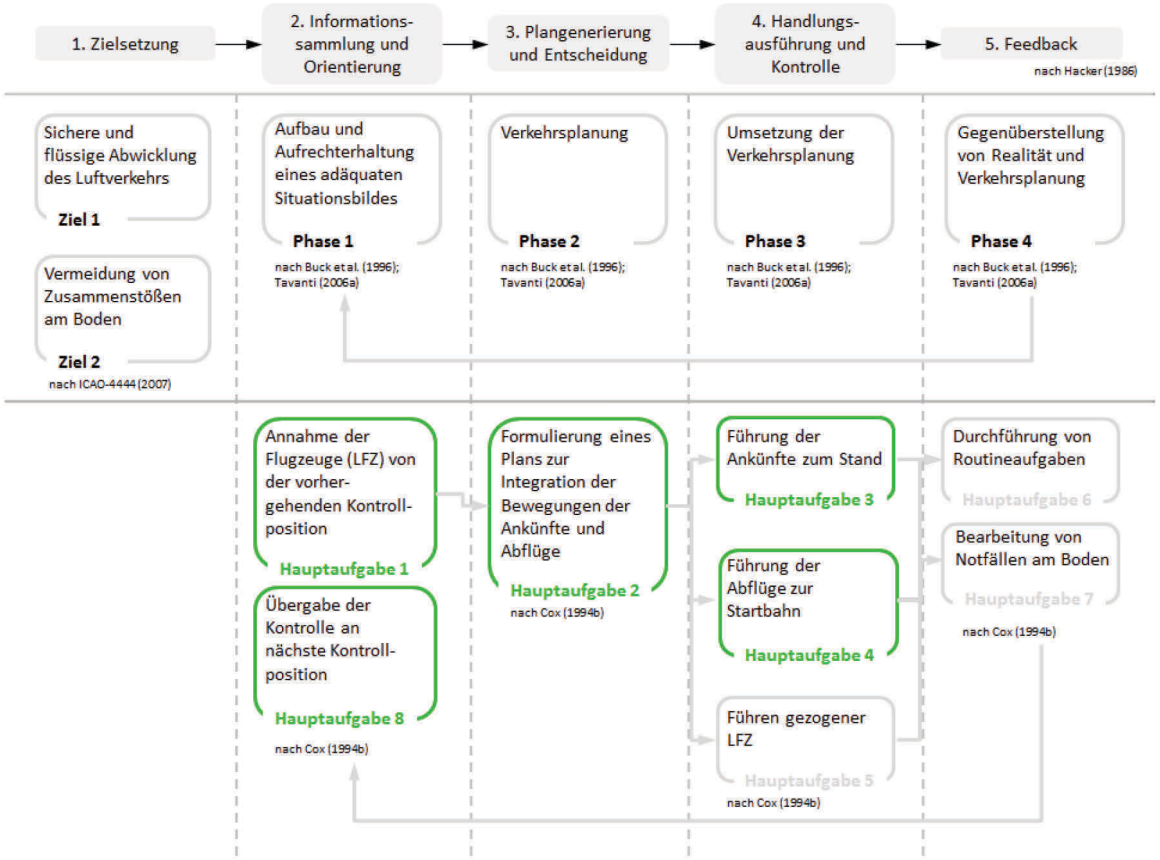

Abbildung 11: Zuordnung der acht Hauptaufgaben des Rollverkehrslotsen (TWR) nach Cox (1994b) zu den Phasen der Handlungsregulation nach Hacker (1986) (nach Smieszek, 2011)

Ferner beinhaltet die Übergabe an die nächste Kontrollposition nur einen Arbeitsschritt und wurde daher in die jeweils vorgelagerten Aufgaben integriert. Die Hauptaufgabe vier „Führung der Abflüge zur Startbahn“ nach Cox (1994b) wurde aufgrund ihrer Komplexität nochmals in zwei Unteraufgaben unterteilt. Hierbei ist zunächst die Aufgabe „Initiieren des Push-Back“ und im Anschluss „Führung der Abflüge zur Startbahn“ auszuführen (vgl. Smieszek, 2011). Für *den Rollverkehrslotsen (GND)* wurden die folgenden vier Hauptaufgaben weiter betrachtet (vgl. Smieszek, 2011):

1. Formulierung eines Plans zur Integration der Bewegungen der Ankünfte und Abflüge
2. Führung der Ankünfte zum Stand,
3. Initiieren des Push-Back,
4. Führung der Abflüge zur Startbahn.

Abbildung 12 auf Seite 75 stellt beispielhaft den Grafen der Hauptaufgabe zwei „Führen der Ankünfte zum Stand“ dar, der die nötigen Aktionen beschreibt, welche der Lotse ausführen muss, um die Unteraufgabe zu bearbeiten: Nachdem der Pilot des betreffenden Flugzeugs das Verlassen der Landebahn bestätigt hat, muss zunächst auf Grundlage der in Hauptaufgabe eins getätigten Beobachtungen und Planungen geprüft werden, ob die Verkehrssituation es zulässt, die gesamte Route freizugeben. Ist dies der Fall, kann die Übermittlung der Freigabe erfolgen. Der Pilot bestätigt daraufhin den Erhalt der Freigabe durch einen Readback[27], welcher vom Lotsen zu überwachen ist. Ist eine Übermittlung der gesamten Route auf Grund der Verkehrssituation nicht möglich, so muss diese Notwendigkeit zunächst erkannt und anschließend überprüft werden, welcher Teil der Gesamtroute sicher ist. Dann erst kann der erste Teil der Freigabe übermittelt werden, wobei wiederum der Readback des Piloten überprüft werden muss. Während das LFZ die Route abrollt, werden dessen Position und Richtung vom Lotsen beobachtet und die Routen der übrigen Flugzeuge mental projiziert. So können mögliche Konflikte vorhergesehen und der Plan entsprechend angepasst werden, bevor der nächste Teil der Route übermittelt wird. Dieser Prozess der Informationssammlung bis zur Übermittlung der neuen Route erfolgt solange, bis das Flugzeug den Stand erreicht hat und die Hauptaufgabe damit erfüllt ist.

Für den *Towerlosen (TWR)* wurden nach Cox (1994b) ebenfalls sieben Hauptaufgaben berichtet:

1. Formulierung eines Plans zur Integration der Bewegungen der Ankünfte und Abflüge
2. Führung der Abflüge
3. Führung der Ankünfte
4. Durchführung von Routineaufgaben
5. Übergabe der Kontrolle über die Abflüge an die nächste Behörde
6. Übergabe der Kontrolle über die Ankünfte an den Rollverkehrslotsen
7. Handlungen nachdem Flugzeug Zuständigkeitsbereich verlassen hat.

Hierbei sind sowohl die Routineaufgaben (Aufgabe 4) und die Handlungen nach Verlassen des Zuständigkeitsbereiches (Aufgabe 7) nicht mit berücksichtigt worden. Ferner wurde, wie oben beschrieben, die Übergabe an die nächste Kontrollposition wiederum in die jeweils vorgelagerte Hauptaufgabe integriert, so dass drei Hauptaufgaben detailliert betrachtet wurden:

[27] Als Readback wird die Wiederholung der Anweisung des Lotsen durch den Piloten bezeichnet. Sie dient der Bestätigung, dass der Pilot die Anweisung des Lotsen exakt verstanden hat.

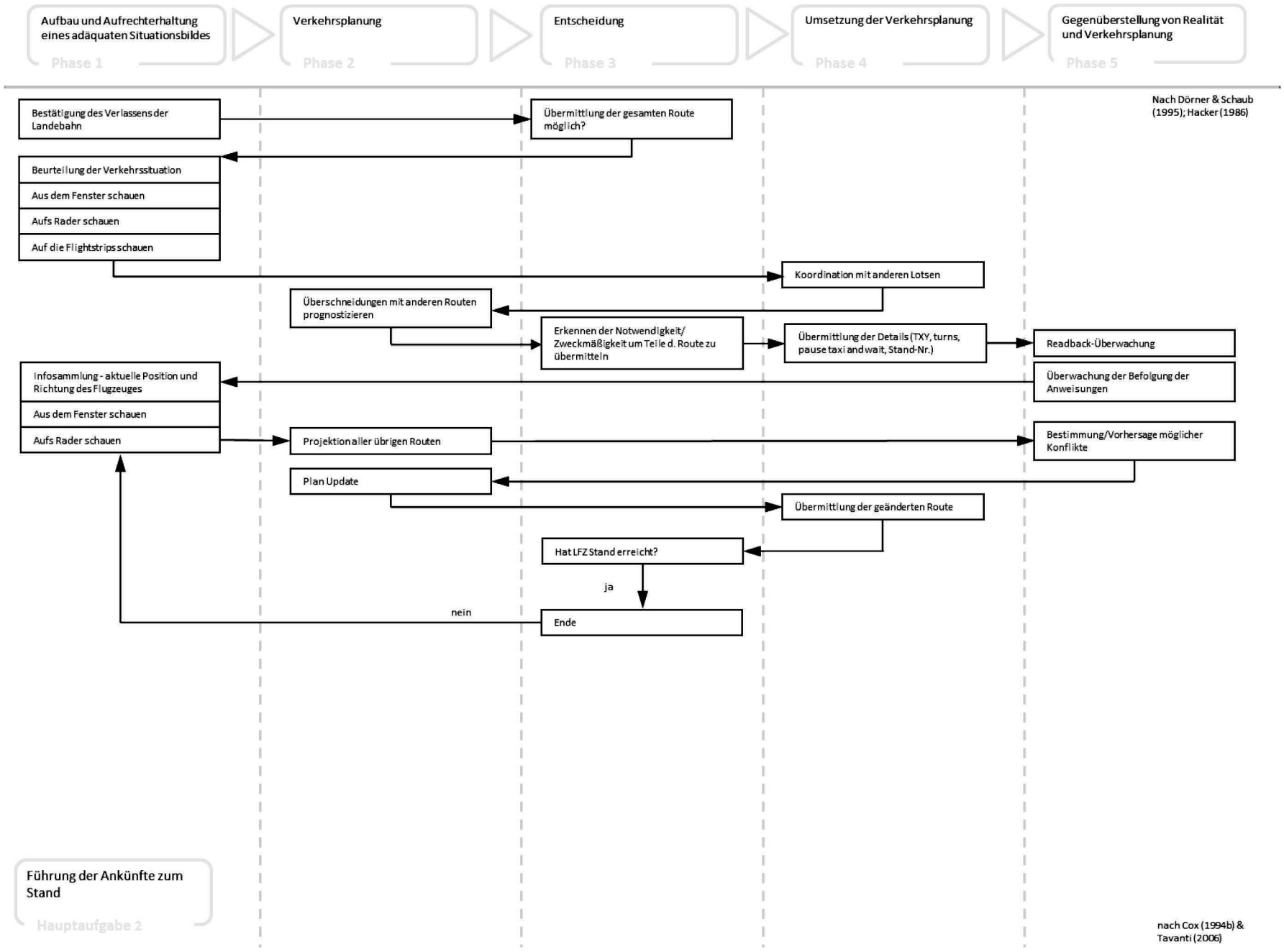

Abbildung 12: Unteraufgaben zur Ausführung der Hauptaufgabe 3 des Rollverkehrslotsen "Führen der Ankünfte zum Stand" (vgl. Smieszek, 2011)

1. Formulierung eines Plans zur Integration der Bewegungen der Ankünfte und Abflüge
2. Führung der Abflüge
3. Führung der Ankünfte

Die Grafen aller Hauptaufgaben sind im Anhang beigefügt.

Auf diese Weise wurden die Ergebnisse der vier in Kapitel 4.3 vorgestellten Aufgabenanalysen zusammengefasst und ein prototypischer Arbeitsablauf der beiden Lotsenpositionen geschaffen, sowie ein Einblick in die zur Aufgabenausführung nötigen kognitiven Prozesse gegeben. Diese prototypischen Arbeitsabläufe dienen als Richtlinie für die Implementierung eines ablauffähigen Modells, welches in Kapitel 5 ausführlich beschrieben wird. da nicht alle Unteraufgaben im gleichen Detailgrad, wie in der Aufgabenanalyse beschrieben in die Modellbildung einfließen können, müssen an verschiedenen Stellen Vereinfachungen vorgenommen, sowie zur Beschreibung der zur Aufgabenausführung nötigen kognitiven Prozesse kognitions-psychologische Theorien umgesetzt werden (vgl. Kapitel 5). Im folgenden Kapitel wird zunächst auf das Beschreibungsmittel eingegangen, welches zur Modellimplementierung verwendet wurde.

4.6 Auswahl des Beschreibungsmittels

Allgemein steht für die modellhafte Beschreibung eine Vielzahl von Beschreibungsmitteln zur Verfügung, da die Modellierung menschlichen Verhaltens hochgradig interdisziplinär erfolgt und sowohl aus dem Methodenrepertoire der Kognitionswissenschaft und Psychologie, als auch aus den Ingenieurwissenschaften geschöpft werden kann (vgl. Kapitel 3.2.2). Ruckdeschel (1997) erstellt eine recht ausführliche Zusammenfassung einschlägiger Methoden verschiedenster Disziplinen (z.B. Logik, Regelungstheorie, Modellierung verteilter Systeme), die zur Modellierung menschlichen Verhaltens in Mensch-Maschine-Systemen verwendet werden können und vergleicht diese hinsichtlich folgender Merkmale:

- Linearität im Sinne der Regelungstheorie: quasilinear/nichtlinear
- Zeitbegriff im Sinne der Regelungstheorie: zeitdiskret/zeitkontinuierlich
- Modellierungsmächtigkeit im Sinne verteilter Systeme: sequentiell/parallel (nebenläufig)
- Klassifikation des Wissens im Sinne wissensbasierter Systeme
- Möglichkeit, die drei Verhaltensarten nach Rasmussen (1983) abzubilden (fertigkeits-, regel- und wissensbasiertes Verhalten)

Der Anspruch einer ganzheitlichen makrokognitiven Beschreibung des Mensch-Maschine-Systems Flughafenverkehrskontrolle erweist sich zur Eingrenzung der Methodenvielfalt dahingehend als nützlich, da mikrokognitive Methoden der kognitiven Modellierung, vor allem kognitive Architekturen und Produktionensysteme, diesem Anspruch nicht, oder zumindest nur

beschränkt gerecht werden können. Vor allem die Schwierigkeit kognitive Architekturen an Simulationen der Außenwelt anzukoppeln, die in der Lage sind, die Aufgaben von Operateuren in komplexen Mensch-Maschine-Systemen adäquat abzubilden und die Prozesse nicht nur sehr abstrahiert darstellen (z.B. bei der Verwendung von Mikrowelten) stellt ein großes Problem dar (vgl. Kapitel 3.2).

Da diese Anbindung allerdings ein zentrales Ziel der hier beschriebenen Modellierung und Teil der Problemstellung ist, sollte ein „ideales" Beschreibungsmittel in der Lage sein, sowohl die menschliche Kognition, als auch den zu kontrollierenden Prozess gleichermaßen beschreiben zu können, sie also in einer „Sprache" abbilden zu können, um aufwändige Schnittstellen-implementierung zwischen Lotsen- und Flughafenmodell zu vermeiden. Daraus ergibt sich die Notwendigkeit, dass ein solches Beschreibungsmittel die Nebenläufigkeit von Prozessen, also das gleichzeitige Ablaufen mehrerer Prozesse, adäquat abbilden kann, da Mensch und Flughafen teilweise unabhängig voneinander agieren und sich nur zeitweise synchronisieren (vgl. Kapitel 4.2). Diese Eigenschaft ist nach Ruckdeschel (1997) vor allem bei Methoden zur Modellierung verteilter Systeme und hier vor allem unter Verwendung von Petrinetzen realisierbar.

Petrinetze, entwickelt von Petri (1962), dienen zur Erstellung von Modellen nebenläufiger Systeme und werden hauptsächlich in der Informatik verwendet. Sie eignen sich allerdings zur Modellierung und Analyse jeglicher Systeme, bei denen Nebenläufigkeit und Kommunikation Schlüsselsrollen einnehmen (Jensen & Kristensen, 2009). Hierbei kann ein Petrinetz als „*a directed graph with two kinds of nodes, interpreted as places and transitions, such that no arc connects two nodes of the same kind*" (Desel & Juhás, 2001, S. 5) definiert werden. Eine Erweiterung klassischer Petrinetze stellen farbige Petrinetze (Coloured Petri Nets, CPN; definiert bei Jensen, 1997) dar. Diese Erweiterung dient der Verbindung der grafischen Modellierung von Systemen einfacher Petrinetze (Petri, 1962; Reisig, 1982) mit den Möglichkeiten einer höheren Programmiersprache (Jensen & Kristensen, 2009). Bei farbigen Petrinetzen handelt es sich (wie auch bei klassischen Petrinetzen) um ein ereignisdiskretes Modellierungswerkzeug, welches eine grafische Darstellung von Prozessen ermöglicht, wobei ein Netz generell aus zwei verschiedenen Gruppen von Elementen (sogenanntem Knoten) besteht: Plätze und Transitionen (werden im Petrinetz als Kreise bzw. Vierecke dargestellt).

Ein System ist dabei als eine Menge von Zuständen oder Bedingungen und Zustandsübergängen definiert und wird über die beiden grafischen Elemente abgebildet (vgl. Abbildung 13). Plätze (dargestellt als Kreise oder Ovale; mit P1 und P2 bezeichnet) stellen dabei Zustände dar, die das System annehmen kann. Diese werden über Kanten – gerichtete Pfeile, über die gedanklich abstrakte Beziehungen zwischen den Komponenten dargestellt werden – mit Transitionen (Rechtecke; mit T1 bezeichnet) verbunden, welche die Zustandsübergänge darstellen. Dabei sind Plätze passive Elemente des Netzes und dienen dem Speichern von

Informationen. Diese Informationen werden als sogenannte Marken (Token; als schwarzer Punkt auf P1 dargestellt) auf den entsprechenden Plätzen abgelegt und beschreiben dabei die Erfüllung oder Nichterfüllung bestimmter, durch den Platz modellierter Bedingungen. Plätze, die einer Transition vorgeschaltet sind werden als Vorplätze (P1), der Transition nachgeschaltete Plätze als Nachplätze (P2) bezeichnet. Pfeile, die zu einem Knoten hin gerichtet sind, werden als Input-Pfeile (Input-Arcs), von einem Knoten weg gerichtete Pfeile als Output-Pfeile (Output-Arcs) bezeichnet.

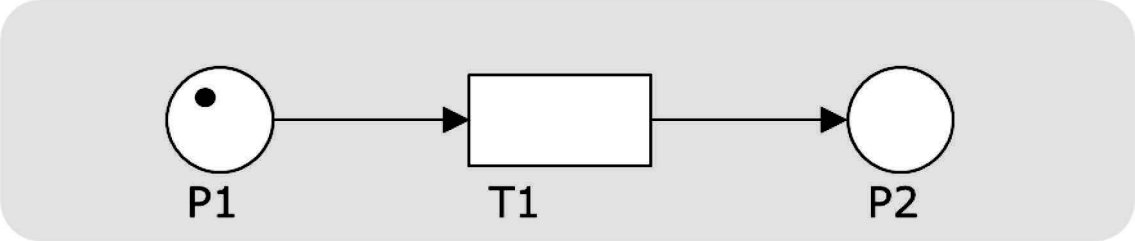

Abbildung 13: Beispiel für ein einfaches Petrinetz

Als aktive Elemente des Netzes sind Transitionen in der Lage Marken auf direkt mit ihnen verbundenen Plätzen durch schalten oder „feuern“ zu verändern. Das Vorhandensein einer Marke beschreibt hierbei das Erfüllt sein einer durch P1 modellierten Bedingung. Durch das Schalten einer Transition werden Marken auf den Vorplätzen konsumiert und auf den Nachplätzen der Transition erzeugt. Bei klassischen Petrinetzen ist es dabei nur möglich eine einzige Art von Marken (sogenannte „schwarze Marken“) zu verwenden. Es können lediglich mehrere gleiche Marken dargestellt werden, indem auf einem Platz mehrere schwarze Punkte eingetragen werden, oder, bei einer sehr großen Anzahl Marken, einfach die Anzahl der Marken als Zahl auf den Platz geschrieben wird. Für die Modellierung komplexer Systeme stellt dies eine starke Einschränkung dar.

Die Entwicklung farbiger Petrinetze ermöglichte es, unterschiedliche Arten von Marken, also Marken mit unterschiedlichem Dateninhalt und Eigenschaften zu verwenden (unterschiedlich farbige Marken[28]). Die funktionale Programmiersprache CPN ML, die auf der Sprache Standard ML aufbaut, stellt dabei die nötigen Routinen zur Verfügung, um Datentypen zu definieren, Datenmanipulation zu beschreiben und kompakte parametrisierbare Modelle zu erstellen (Jensen & Kristensen, 2009). Hierdurch wird es ermöglicht, dass eine Marke beispielsweise ein Flugzeug (mit Typ, Rufzeichen, Ziel usw.) repräsentieren kann (Manske, Smieszek, Hasselberg, & Möhlenbrink, 2013).

Die Abbildung 14 stellt ein einfaches farbiges Petrinetz dar, welches den Landeanflug eines Flugzeuges beschreibt. Hierbei sind die Plätze „Final“ und „Landed“ jeweils als Zustände des Flugzeuges vor (im Landeanflug: Final) bzw. nach (gelandet: Landed) der Landung anzusehen.

[28] Die Farbe einer Marke beschreibt ihren Dateninhalt und steht in keinerlei Zusammenhang zur farblichen Gestaltung von Netzen in CPN-Tools.

Die Transition „Landing“ beschreibt den Prozess des Landens selbst. Das betreffende Flugzeug ist als Marke modelliert, die in der Abbildung 14 als grüner Punkt dargestellt ist. Es ist zu erkennen, dass Marken nicht mehr nur als schwarze Punkte dargestellt werden, sondern einerseits durch die Ziffer „1“ dargestellt wird, wie viele gleichartige Marken vorhanden sind, und andererseits, welchen Dateninhalt die jeweilige Marke besitzt (im Fall des Flugzeuges das Rufzeichen „BER123“). Ein Colorset ist vergleichbar mit den Variablen einer normalen Programmiersprache.

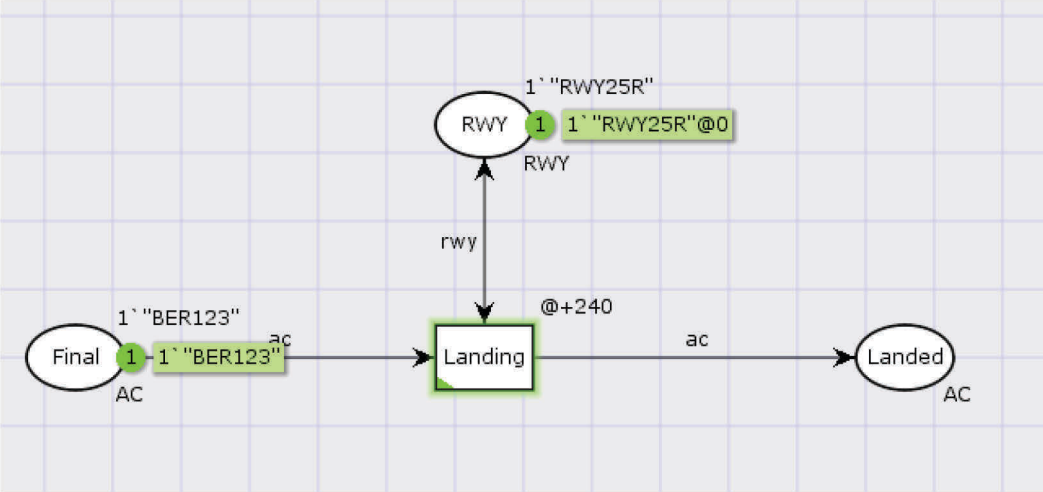

Abbildung 14: Beispiel für ein einfaches farbiges Petrinetz für einen Landeprozess

Auf den Plätzen des Colorsets „AC“ können somit nur Marken präsent sein, die genau von diesem Colorset sind. Eine weitere Marke ist auf dem Platz „RWY“ zu erkennen. Diese beschreibt das Vorhandensein einer Start-/ Landebahn mit der Bezeichnung „RWY25R“ und verwendet das Colorset „RWY“. Diese Marke besitzt weiterhin die Besonderheit, dass sie Zeitinformationen beinhaltet, was an dem angefügten „@0“ zu erkennen ist. Diese Symbolik gibt Aufschluss darüber, zu welchem Modellzeitpunkt die Marke für die angeschlossene Transition zur Verfügung steht. Solange der Wert der Modellzeit kleiner als der hier angegebene Zeitwert ist, gilt die Marke als noch nicht verfügbar und steht nicht für die Transition zur Verfügung.

Durch das Feuern der Transition „Landing“, wird der Ablauf eines Landeprozesses modelliert und die Verteilung der Marken im Netz (die sogenannte Markierung) verändert. Dabei werden die betreffenden Marken auf den Vorplätzen verbraucht und auf den Nachplätzen die jeweiligen Marken generiert. Hat die Transition gefeuert, wird ein neuer Systemzustand erreicht. Abbildung 15 stellt das obige Petrinetz nach Auftreten der Transition „Landing“ dar. Es ist zu erkennen, dass die Marke, die das Flugzeug modelliert, vom Platz „Final“ auf den Platz „Landed“ versetzt worden ist (das Flugzeug ist gelandet).

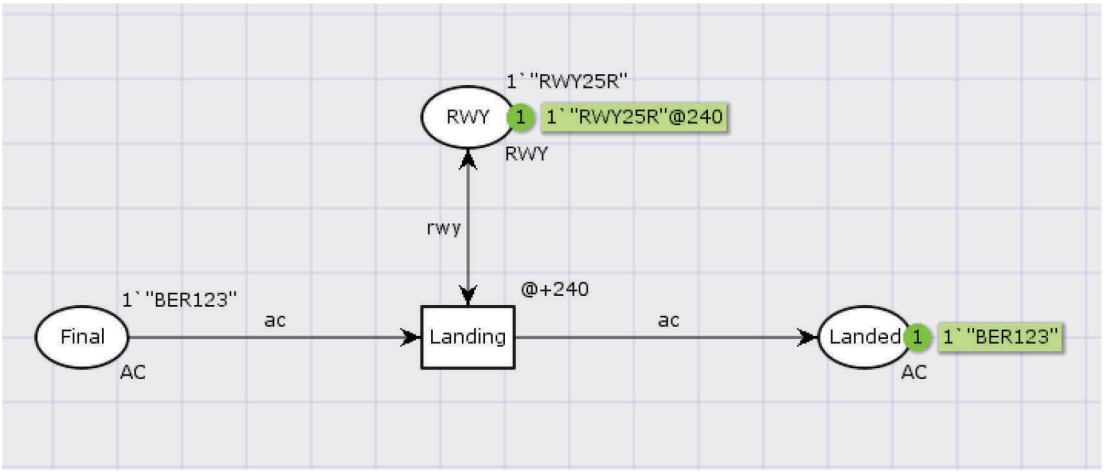

Abbildung 15: Das Petrinetz aus Abbildung 14 nach einmaligem feuern der Transition "Landing"

Gleichermaßen wurde die Zeit der Start-/Landebahn heraufgesetzt und zwar um den Betrag, der an der Transition „Landing“ angegeben ist (im Beispiel 240 Zeitschritte, dargestellt durch „@+240“an der Transition). Es wurde damit festgelegt, dass bei einer Landung die Start-/ Landebahn für 240 Zeitschritte nicht von anderen Flugzeugen verwendet werden darf. Die Marke „RWY“ steht demnach 240 Zeitschritte nicht für eine andere Landung zur Verfügung.

Da Modelle komplexer dynamischer Systeme sehr schnell unübersichtlich und verwirrend werden können, wurde das Konzept der hierarchischen Petrinetze eingeführt. Komplexe CPN-Modelle können dabei in Sub-Module unterteilt werden, um dem Modellierer eine Arbeit auf unterschiedlichen Abstraktionsebenen zu ermöglichen (Jensen & Kristensen, 2009). Mit Hilfe dieser hierarchischen Struktur ist es möglich, ein großes Modell zu erstellen, indem eine Vielzahl kleinerer Modelle kombiniert wird. Um Sub-Module darzustellen, werden sogenannte Substitutions-Transitionen eingeführt, welche als Rechteck mit doppeltem Rahmen dargestellt werden.

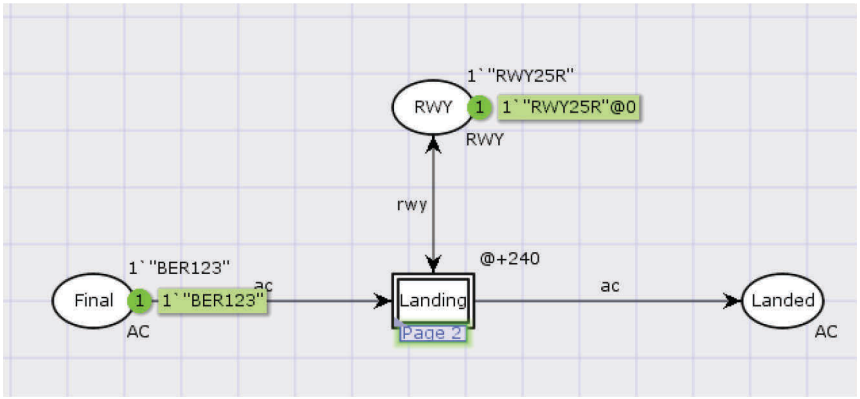

Abbildung 16: Beispiel für eine Substitutions-Transition in einem hierarchischen Petrinetz

In Abbildung 16 ist dies beispielhaft für die Transition „Landing“ umgesetzt. Auf der gezeigten Ebene wird nicht dargestellt, was exakt bei der Landung passiert. Jedoch kann bei Bedarf die

hierarchisch darunterliegende Ebene (auch Seite genannt) aufgerufen werden, welche den Prozess der Landung im Detail beschreibt[29].

Zusätzlich zu der bereits besprochenen Möglichkeit, die Nebenläufigkeit des Flughafenverkehrskontrollsystems adäquat abzubilden, bieten farbige Petrinetze eine Reihe weiterer Vorteile: So kann mit dieser Modellierungsmethode die angestrebte Beschreibung aller drei Modellteile (Flughafen-, Interaktions- und Lotsenmodell) gewährleistet werden. Farbige Petrinetze sind darüber hinaus mathematisch definiert und analysierbar, für die Anwendung ist aber lediglich ein Verständnis der Syntax und Semantik notwendig. Die Modularisierbarkeit und hierarchische Unterteilung von Modellen bietet zusätzlich eine gute Handhabung der Komplexität des zu modellierenden Prozesses und ermöglicht die individuelle Erweiterung und damit die Wiederverwendbarkeit von Modellen. Dem kommt auch die relative Stabilität von Modellen gegenüber kleineren Änderungen entgegen (Jensen, 1997a), so dass Erweiterungen oder Änderungen an Modellen relativ schnell und ohne viel Aufwand möglich sind. Farbige Petrinetze bieten zudem die Möglichkeit der Einbindung von Zeit (wie oben beschrieben), um so Abhängigkeiten in der Zeit beschreiben zu können, die für die Flughafenverkehrskontrolle von essentieller Bedeutung sind. Zusätzlich bietet die relativ intuitive grafische Darstellung den Vorteil einer anschaulichen Beschreibung und einer einfachen Nachvollziehbarkeit und „Sichtbarmachung“ von Prozessen, was die Transparenz und den wissenschaftlichen Austausch der erstellten Modelle erleichtert.

Nicht unerheblich ist ebenso die Tatsache, dass Petrinetze und vor allem farbige Petrinetze bereits mehrfach erfolgreich angewendet wurden, um verschiedene Problemstellungen in der Luftfahrt (Davidrajuh & Lin, 2011; Kovács, Németh, & Hangos, 2005; Lesire & Tessier, 2005; Oberheid & Söffker, 2008; Oberheid, 2006; Stütz, 2000; Vidosavljević & Tošić, 2010), sowie menschliches Verhalten (Möhlenbrink & Friedrich, 2009; Möhlenbrink et al., 2012; Möhlenbrink, 2011; Ruckdeschel & Onken, 1994; Ruckdeschel, 1997; Werther et al., 2007; Werther, 2006a, 2006b) zu modellieren und zu analysieren. Nichtsdestotrotz ist die Anwendung farbiger Petrinetze zur Beschreibung menschlichen Verhaltens bisher auf die Erstellung formaler Modelle beschränkt. Die Modellierung kognitiver Funktionen stellt damit einen innovativen Ansatz dar.

[29] Eine ausführliche Einführung in Petrinetze kann beispielsweise bei Baumgarten (1996) oder Reisig (2010) gefunden werden. Eine Kurzdarstellung des Konzeptes beschreiben Desel und Juhás (2001). Eine ausführliche Einführung in farbige Petrinetze geben Jensen (1997a, 1997b, 1997c), Jensen (1998), sowie Jensen und Kristensen (2009).

Nicht zuletzt ist auch die Tatsache, dass einschlägige Computertools vorhanden sind, um die Modellierung, Simulation und Analyse von Petrinetzmodellen zu gewährleisten, ein ausschlaggebender Punkt für deren Verwendung. Eines der bekanntesten und mächtigsten Werkzeuge stellt hier zweifelsohne CPN-Tools dar, welches auch zur Erstellung des vorliegenden Modells verwendet wurde (CPN-Tools, o.J.; Jensen, Kristensen, & Wells, 2007; Ratzer et al., 2003). Für die Erstellung des Modells und zur Lösung der beschriebenen Problemstellung fiel die Entscheidung aus folgenden Gründen zugunsten der Verwendung farbiger Petrinetze als Beschreibungsmittel aus:

- Möglichkeit sowohl den Flughafenprozess als auch die kognitiven Prozesse und das Verhalten des Lotsen in einer „Sprache“ abzubilden
- Keine Schnittstellenimplementierung zwischen kognitiver Simulation und dem zu kontrollierenden Prozess
- Berücksichtigung der Nebenläufigkeit von Prozessen
- Gute Handhabung der Komplexität und individuelle Erweiterbarkeit durch Modularisierbarkeit und hierarchische Strukturierung
- Relative Stabilität von Modellen gegenüber kleineren Änderungen
- Möglichkeit der Einbindung von Zeit
- Intuitive Darstellung und damit Nachvollziehbarkeit und Transparenz
- Vorhandensein einschlägiger Computertools zur Modellierung, Simulation und Analyse von Petrinetzmodellen
- Mehrfache erfolgreiche Verwendung zur Modellierung von Problemstellungen in der Luftfahrt, Flugverkehrskontrolle sowie der Beschreibung kognitiver Prozesse und menschlichen Verhaltens.

4.7 Kapitelzusammenfassung

Bei der Durchführung der theoretischen Vorarbeiten zur Entwicklung eines makrokognitiven Modells mentaler Beanspruchung von Towerlotsen dienten die sechs von Cacciabue (1998b) definierten Schritte für die Entwicklung kognitiver Simulationen als Richtlinie. Hierbei wurde argumentiert, dass die bestehenden Richtlinien um einen siebten, essentiellen Schritt der Modellbildung ergänzt werden müssen: die Validierung, also der Verglich von Modellverhalten mit experimentellen Daten. Es ist die Aufgabe des Modellierers, über die Validierung den Bezug zwischen dem Modell und dem realen Untersuchungsgegenstand herzustellen (Möhlenbrink, 2011; vgl. auch Bub & Lugner, 1992) (vgl. Kapitel 4.1). Die ersten fünf Schritte der Cacciabue-Richtlinien können als theoretische Vorarbeiten gesehen werden und wurden im vorangegangenen Kapitel näher erläutert: So wurden zunächst als *Simulationsziel* die modellhafte Beschreibung und Quantifizierung mentaler Beanspruchung definiert. Um einen

Rahmen für die Modellerstellung zu schaffen und nötige Vereinfachungen zu definieren, wurden *Problemgrenzen* festgelegt: So wird eine Dreiteilung des Modells in Flughafen-, Interaktions- und Lotsenmodell abgestrebt, um eine ganzheitliche Beschreibung des Mensch-Maschine-Systems und der mentalen Beanspruchung gewährleisten zu können. Dabei werden aufbauend auf den Arbeiten von Koros et al. (2003) diejenigen Einflussfaktoren berücksichtigt, die Lotsen als am wichtigsten bewerten. Vor allem wird der Einfluss des Verkehrsaufkommens auf die mentale Beanspruchung des Lotsen abgebildet (vgl. Kapitel 4.2). Zur Erlangung eines intensiven Verständnisses der Aufgaben und des Systems wurden anschließend fünf bestehende *Aufgabenanalysen* hinsichtlich ihrer Verwendbarkeit zur Erstellung eines Aufgabenmodells überprüft und zusammengefasst (vgl. Kapitel 4.3). Hierbei wurden Haupt- und Unteraufgaben der beiden Positionen TWR und GND identifiziert, wobei sich die Aufgabenanalyse von Cox (1994a, 1994b) als sehr detailliert erwies, sodass diese vorrangig für die Erstellung des theoretischen Models verwendet wurde. Im nächsten Schritt wurde das System Flughafenverkehrskontrolle sowohl durch eigene *Feldbeobachtungen* als auch durch *Analyse von Vorschriftenwerken und Regularien* analysiert und somit bestehende Regeln der Flughafenverkehrskontrolle, die Zuständigkeiten der Lotsen, der Aufbau des Lotsenarbeitsplatzes, sowie wichtige Hilfsmittel analysiert. Als wichtigste Informationsquellen konnten die Außensicht, die Funkkommunikation, sowie die Flugstreifen identifiziert werden, die auch in der Modellierung Berücksichtigung finden sollen. Ferner werden in der Praxis zwei Positionen (TWR und GND) mit unterschiedlichen Zuständigkeiten eingesetzt, wovon im vorliegenden Modell aus Gründen der Einfachheit nur ein Lotsenmodell erstellt werden wird, welches für beide Positionen zuständig ist. (vgl. Kapitel 4.4). Die so erlangten Erkenntnisse wurden in ein *theoretisches Modell* implementiert, welches prototypische Arbeitsabläufe von Rollverkehrs- und Towerlotsen anhand der Phasen der Handlungsregulation nach Hacker (1986) systematisiert, welche einerseits die nötigen auszuführenden Handlungen und andererseits die zur Handlungsausführung nötigen kognitiven Prozesse beschreiben (vgl. Kapitel 4.5). Als geeignetes *Beschreibungsmittel*, welches in der Lage ist, die im Kapitel 4.2 definierten Ziele und Problemgrenzen hinreichend zu beschreiben, wurden farbige Petrinetze ausgewählt. Diese bringen insbesondere die nötige Flexibilität mit sich, um die Komplexität des Systems zu beherrschen und die drei Modellteile in einer „Sprache“ abbilden zu können. Darüber hinaus sind einschlägige Computertools zur Modellierung und Analyse vorhanden. Es wurde weiterhin eine grundlegende Einführung in farbige Petrinetze gegeben (vgl. Kapitel 4.6). Mit den hier getätigten Analysen wurden die ersten fünf Schritte des Cacciabue-Frameworks bearbeitet. Die Schritte sechs und sieben werden in den folgenden Kapiteln (5 und 6) behandelt.

5 Beschreibung und Implementierung des entwickelten Modells

Das vorliegende Kapitel beschreibt die Umsetzung der in Kapitel 4 vorgestellten theoretischen Vorarbeiten in eine lauffähige kognitive Simulation. Hierzu werden im Abschnitt 5.1 zunächst allgemeine Anmerkungen zum Modell gemacht. Im Kapitel 5.2 wird anschließend das Flughafenmodell zunächst theoretisch, anschließend dessen Implementierung im Petrinetz genauer erläutert. Kapitel 5.3 beschäftigt sich auf gleiche Weise mit dem Interaktionsmodell. Fokus des Kapitels 5.4 liegt auf der Implementierung des Lotsenmodells, wobei zu den einzelnen Sub-Modulen `perceive` (Kapitel 5.4.2 und 5.4.3), `plan` (Kapitel 5.4.4 und 5.4.5), `decide` (Kapitel 5.4.6 und 5.4.7), `act` (Kapitel 5.4.8 und 5.4.9), sowie `Working_Memory` (Kapitel 5.4.10 und 5.4.11) jeweils erst die theoretischen Überlegungen und anschließend die Implementierung im Modell vorgestellt werden. Kapitel 5.5 beschreibt die Simulation und Analyse des Petrinetzmodells, Kapitel 5.6 fasst das Kapitel zusammen.

5.1 Allgemeine Anmerkungen zum Modell

Aufbauend auf den in Kapitel 4 beschriebenen Analysen und theoretischen Vorarbeiten erfolgt die Erstellung des holistischen Flughafen-Lotsen-Modells mit farbigen Petrinetzen als Beschreibungsmittel in der Modellierungs- und Simulationssoftware CPN Tools (Jensen et al., 2007). Die aktuellen Modellversionen (1.8 und 1.9) sind auf der CD im Anhang zu finden. Sie wurde in der CPN Tools-Version 4.0 erstellt, und sind nicht mehr mit den CPN Tools-Versionen 3.4 und darunter kompatibel. Die hier dargelegten Ausführungen gelten für das Modell in seiner Version 1.9, anhand derer auch das Validierungsexperiment durchgeführt wurde. Hierbei wurden im Vergleich zur Version 1.8 einige Modifikationen vorgenommen, um das experimentelle Setting adäquat abzubilden.

Die Modellierung vor allem der physischen Bewegungen des Lotsen erfolgt in Version 1.8 anhand eines fiktiven Lotsenarbeitsplatzes. Dieser wurde in Version 1.9 an das im Versuch verwendete experimentelle Setting angepasst. So war beispielsweise die Entnahme von Flugstreifen aus dem Drucker und der Einsortierung in die Stripbay nicht Bestandteil der Aufgabe der Versuchspersonen. Hierdurch entfallen in Version 1.9 alle damit verbundenen physischen Bewegungen. Es wird an den entsprechenden Stellen nochmals explizit auf die Unterschiede hingewiesen.

Die Erstellung des holistischen Mensch-Maschine-Modells der Flughafenverkehrskontrolle (hier mit MATriCS, Model of Airport TRaffIc Control System bezeichnet) erfolgte anhand der von Jensen (1997a, S.32 ff.) beschriebenen zwölf Schritte zur Erstellung von CPN-Modellen (vgl. Smieszek & Karl, 2012):

1. Beginne damit, die wichtigsten Elemente des zu modellierenden Systems zu identifizieren
2. Beachte den Zweck deines Modells und bestimme einen angemessenen Detailgrad
3. Versuche gute mnemonische Namen für Objekte, Prozesse, Zustände und Aktionen zu finden
4. Versuche nicht alle Aspekte des Systems bereits in der ersten Modellversion abzudecken
5. Wähle einen der Prozesse des zu modellierenden Systems und versuche ein isoliertes Netz für diesen Prozess zu erstellen
6. Benutze die Netzstruktur um Regeln und die Netzinschriften um Datenmanipulation zu modellieren
7. Unterscheide zwischen verschiedenen Arten von Marken
8. Verwende unterschiedliche Arten von Colorsets
9. Erweitere das Prozessnetz durch die Beschreibung der Kommunikation und Interaktion des Prozesses mit anderen Prozessen
10. Ermittele, ob ähnliche Prozesse vorhanden sind
11. Kombiniere die Unternetze der einzelnen Prozesse zu einem großen Modell
12. Die Erstellung eines CPN-Modells ähnelt sehr der Erstellung eines Computerprogrammes.

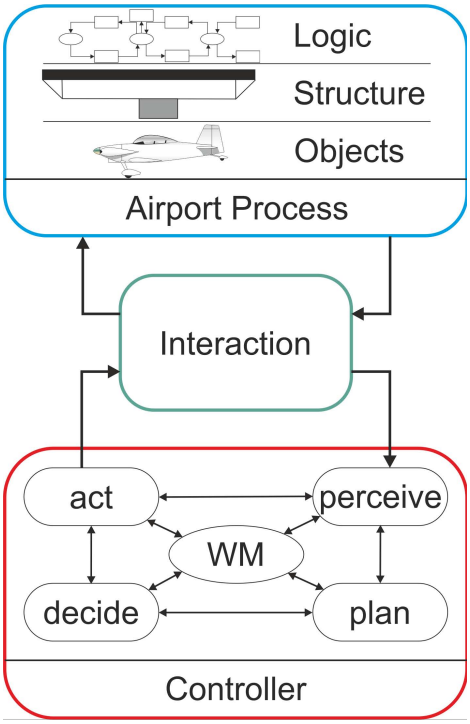

Abbildung 17: Darstellung der drei Teile des Petri-Netz-Modells (Flughafen-, Interaktions- und Lotsenmodell) und deren Verbindungen (nach Smieszek et al., 2013)

Das Modell ist, wie bereits beschrieben, in Anlehnung an Cacciabue (1998b) und Werther (2006a) in drei Hauptmodelle unterteilt: Das Flughafenmodell, das Interaktionsmodell und das Lotsenmodell. Eine schematische Darstellung des Modellaufbaus zeigt Abbildung 17. Die Abbildung 18 auf Seite 86 zeigt die konkrete hierarchische Struktur des Modells mit allen untergeordneten Ebenen. Der obersten Ebene (`Aerodrome_ATC`) direkt untergeordnet sind die Sub-Module `Airport_Process`, `TWR_Controller` und `Interaction` mit ihren jeweiligen Sub-Modulen. Im `Airport_Process` sind die jeweiligen ablaufenden Prozesse als Unterebenen umgesetzt: `Taxiing`, `Enter_RWY`, `Leave_RWY`, `On_Block`, `Off_Block`[30], `Landing`, `Take_off`, `Aerodrome`. Die genaue Bedeutung und Umsetzung wird im Abschnitt 5.2 beschrieben. Das Interaktionsmodell besitzt keine weiteren Unterebenen. Hier wird der Informationsaustausch zwischen Lotsen- und Flughafenmodell über die zur Verfügung stehenden Hilfsmittel dargestellt.

[30] On-Block und Off-Block bezeichnet das auf- und abrollen von den Parkpositionen.

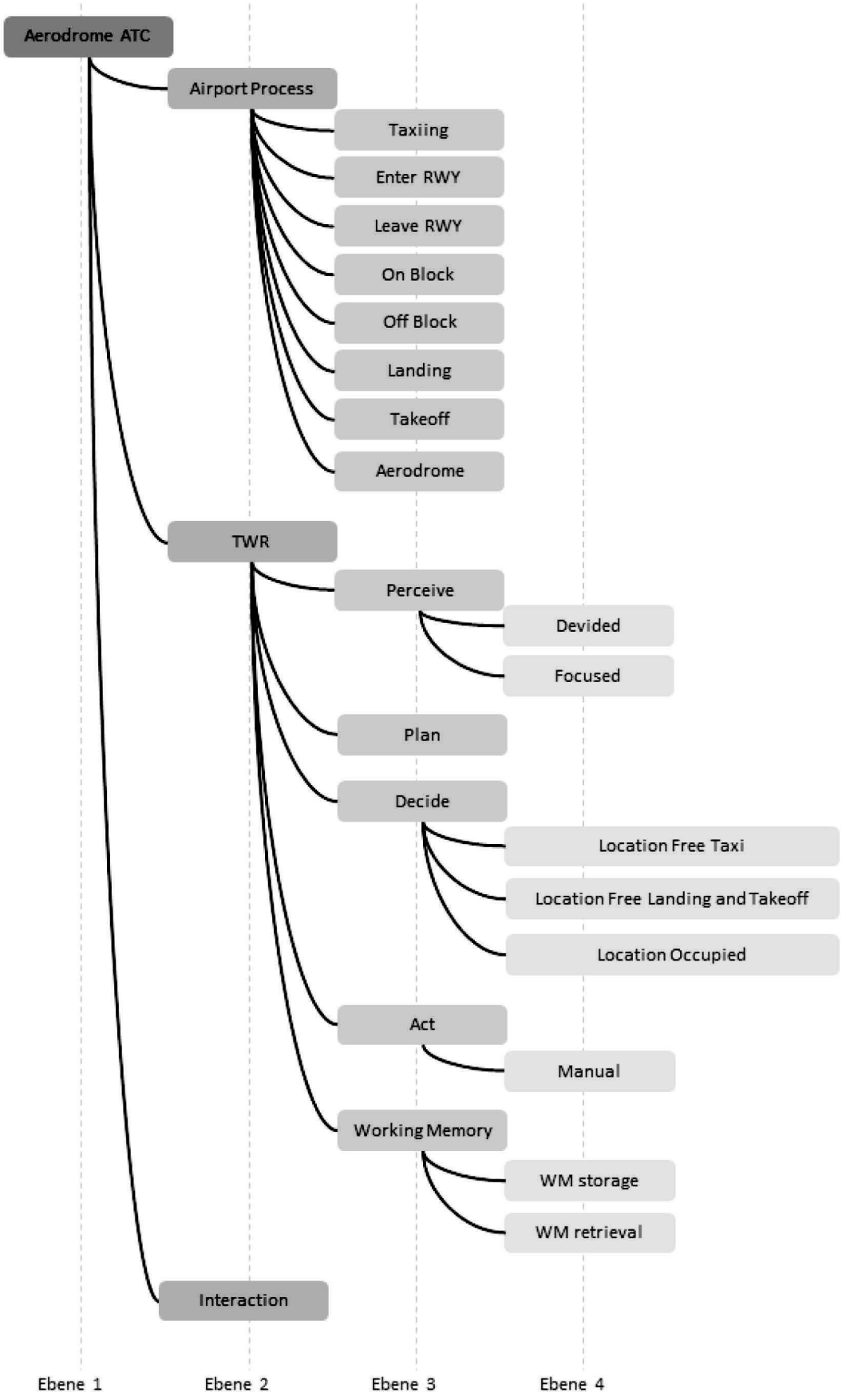

Abbildung 18: Schematische Darstellung der hierarchischen Stuktur des MATriCS-Modells

Eine genauere Erläuterung erfolgt im Kapitel 5.3. Im Lotsenmodel sind die Phasen der Handlungsregulation als Sub-Module umgesetzt: `Perceive`, `Plan`, `Decide`, `Act`. Zusätzlich ist ein Sub-Modul `Working_Memory (WM)` implementiert. Alle drei Hauptteile des Modells sind wichtig für die holistische Abbildung des Flughafenverkehrskontrollsystems; der Fokus liegt allerdings auf der modellhaften Beschreibung der menschlichen Kognition, im Hinblick auf die Replikation mentaler Beanspruchung von Towerlotsen vor allem auf der Komponente Arbeitsgedächtnis.

5.2 Das Flughafenmodell

5.2.1 Theoretische Erläuterungen

Das Flughafenmodell ist ein in Zusammenarbeit mit dem Deutschen Zentrum für Luft- und Raumfahrt (DLR) in Braunschweig entwickeltes generisches Petri-Netz-Modell (vgl. Manske, Smieszek, et al., 2013), welches zunächst beispielhaft für den Flughafen Braunschweig-Wolfsburg (ICAO-Code: EDVE) umgesetzt wurde. Die große Stärke des Modells ist seine einfache Anpassbarkeit zur Modellierung weiterer Flughäfen, was Manske, Smieszek, et al. (2013) anhand des Berliner Flughafens BER (ICAO-Code: EDDB) demonstrieren konnten. Dies ist vor allem wichtig für den durchgeführten Validierungsversuch (vgl. Kapitel 6).

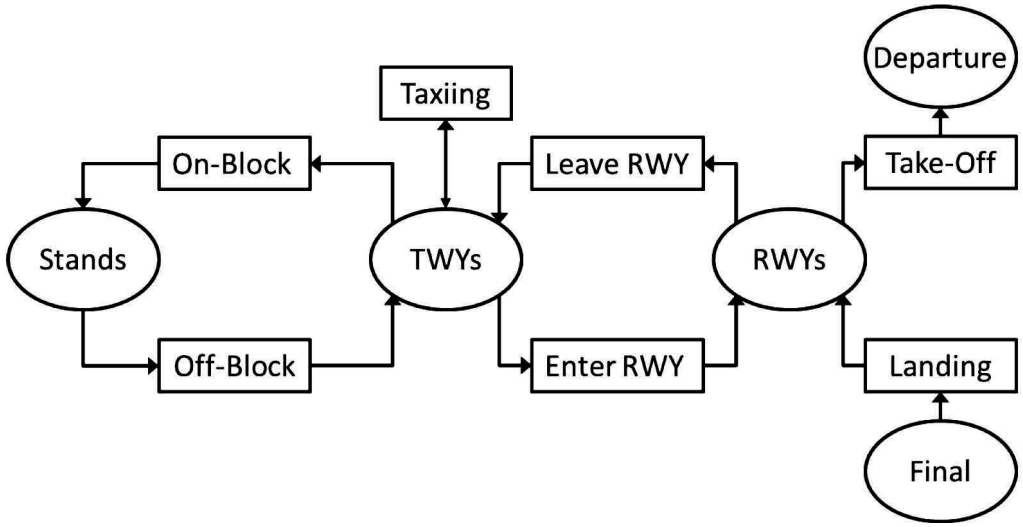

Abbildung 19: Elemente der invariablen Prozesslogik (eigene Darstellung nach Manske, Smieszek, Hasselberg & Möhlenbrink, 2013)

Um diese einfache Anpassbarkeit zu erreichen, wurde die Modellentwicklung auf Basis der Feststellung angegangen, dass einige wenige Orte gleicher Funktionalität an jedem Flugplatz gegeben sind (Manske, 2013). Diese sind: Eine oder mehrere Start- und Landebahnen, An- und Abflugbereiche, Rollwege und Stellplätze für Luft- und andere Fahrzeuge. Im Petri-Netz sind diese allgemeinen Orte daher als Plätze modelliert. Diese werden durch Transitionen miteinander verbunden, welche diejenigen Bewegungen beschreiben, die die auf dem Flughafen befindlichen Objekte ausführen können, um sich von Ort zu Ort zu bewegen. Zusammen bilden

sie die *invariable Prozesslogik* (siehe Abbildung 19). Stark abstrahiert folgt jeder Flugplatz dieser Struktur (Manske, 2013).

Unterschiede, die hinsichtlich des Aussehens und der Komplexität auftreten, werden durch die *variable Prozessstruktur beschrieben*. Die Grundstruktur des Netzes ist so allgemein gehalten, dass ohne eine Veränderung der Netzstruktur verschiedene Flugplätze beschrieben werden können. Die Informationen über die Struktur des jeweiligen Flugplatzes im Besonderen, sind unabhängig von der beschriebenen Grundlogik und werden über Marken eingelesen. Hierin sind die Informationen enthalten, welche Rollwege, Stellplätze und Pisten im Speziellen vorhanden sind, wie diese bezeichnet werden und wie sie orientiert sind (z.B. RWY, A, 26, Std steht zum Beispiel für eine Start-/ Landebahn mit der Bezeichnung A, der Ausrichtung 26, welche in Standardrichtung verwendet wird). Zusätzlich sind auch die Verknüpfungen und Verbindungen der einzelnen Flughafenteile untereinander hinterlegt. Dies sind zum Beispiel die Verbindungen verschiedener Rollwege untereinander (siehe Abbildung 20, Rollweg B ist mit der Piste, sowie mit Rollweg C verbunden, nicht aber mit Rollweg A). Dritter Bestandteil des Flughafenmodells sind die *dynamischen Objekte*. Diese beschreiben welche Luft- und Bodenfahrzeuge sich auf dem Flugplatz an welchen Positionen befinden und welche Freigaben und Intentionen sie besitzen. Im Folgenden wird das Petrinetz-Modell des Flughafens im Detail vorgestellt.

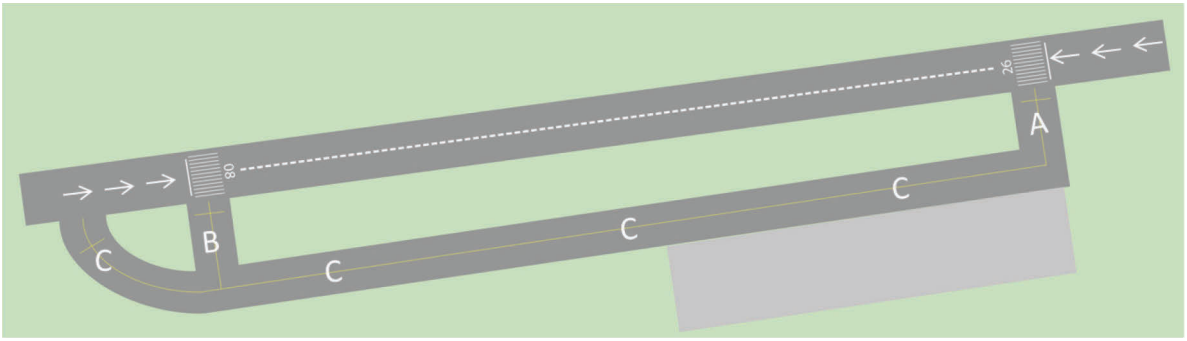

Abbildung 20: Layout eines exemplarischen Flughafens (eigene Darstellung)

5.2.2 Das Petrinetz-Modell im Detail

Die obere Ebene des Modells besteht aus fünf Plätzen und acht Transitionen (in Abbildung 21 auf Seite 89 gelb dargestellt[31]), welche die invariable Prozesslogik darstellen, sowie zwei weiteren Plätzen, welche die variable Prozessstruktur (Location), sowie die dynamischen Objekte (Vehicles) beschreiben (in Abbildung 21 auf Seite 89 blau dargestellt). Die fünf Plätze der invariablen Prozesslogik repräsentieren die allgemeinen Orte des Flughafens: `Stands`[32], Taxiways (`TWYs`), Runways (`RWYs`), welche die Parkpositionen, die Rollwege und

[31] Die farbliche Gestaltung des Netzes dient lediglich der Übersichtlichkeit und hängt nicht mit den Colorsets der einzelnen Plätze zusammen.
[32] „Stand“ bezeichnet die Parkpositionen.

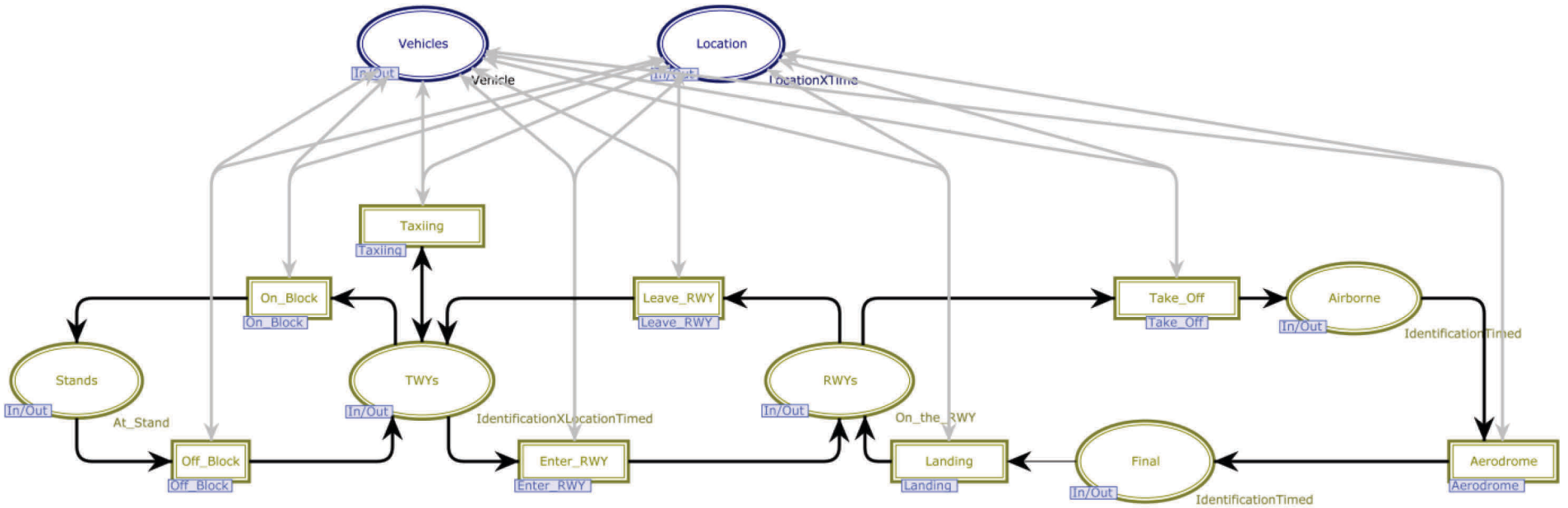

Abbildung 21: Oberste Ebene des Flughafenmodells

die Start-/Landebahnen bezeichnen sowie `Airborne`[33] und `Final`, welche die Umgebung in der Luft um den Flughafen bezeichnen. Auf den Plätzen der invariablen Prozesslogik sind Marken enthalten, welche Repräsentationen derjenigen Objekte (Vehicle) darstellen, die sich entsprechend ihrer Intention und Freigabe über die acht Transitionen bewegen (Manske, 2013). Diese Transitionen beschreiben die Bewegungen, die die auf dem Flughafen befindlichen dynamischen Objekte ausführen können: `On_Block/Off_Block` (rollen an die Parkposition oder von der Parkposition weg), `Leave_RWY/Enter_RWY` (auf- und abrollen auf/von der Start-/Landebahn), `Landing/Takeoff` (Landung und Start), `Aerodrome` (Bewegungen in der Luft). Sie sind jeweils nochmals hierarchisch unterteilt, um die genaue Markenveränderung bei Auftreten einer Aktion beschreiben zu können.

Am Beispiel des Rollens (`Taxiing`) soll die Implementierung der variablen Prozessstruktur erklärt werden: In Abbildung 23 auf Seite 91 sind die beim Rollen verbundenen Elemente dargestellt (vgl. Manske, 2013). Die Transition `Taxiing` sowie der Platz `TWYs` sind Bestandteil der invariablen Prozesslogik. Die Plätze `Location`, `Connections` und `Duration` sind Bestandteil der variablen Prozessstruktur und werden im Folgenden erläutert. Die auf ihnen gespeicherten Marken definieren das Layout des Flughafens (Manske, 2013).

Der Platz `Location` enthält Informationen über alle Teile des Flughafens, also seine Pisten, Rollwege, Stellplätze usw. Um die Bewegung der Luftfahrzeuge über das Rollfeld und die sukzessive Fortbewegung der Luftfahrzeuge im ereignisdiskreten Modellierungssystem des Petrinetzes simulieren zu können, werden die Rollwege weiter in kleinere, möglichst aber gleichgroße, Abschnitte unterteilt. Hierbei wird beispielsweise Rollweg C in C1, C2 usw. unterteilt (vgl. Abbildung 22).

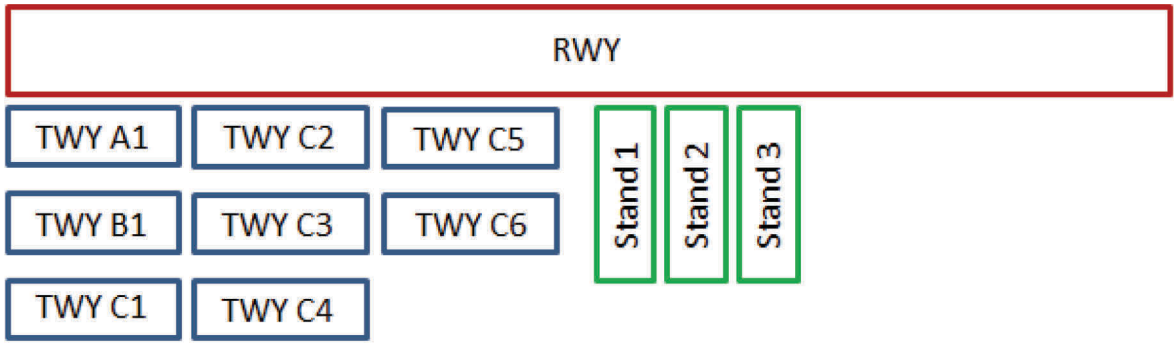

Abbildung 22: Einzelne Teile eines exemplarischen Flughafens (eigene Darstellung nach Manske, 2013)

Auf diesen Abschnitten kann sich jeweils nur ein Objekt zur gleichen Zeit befinden, sodass diese Abschnitte auch als Ressourcen des Flughafens verstanden werden können (Manske, 2013). Sie sind daher für die Zeit, in der ein Objekt diese benutzt, „verbraucht". Die Dauer dieser Belegung wird auf dem Platz `Duration` als Standardzeit für jeden Rollwegsabschnitt festgelegt. Es sollte

[33] „Airborne" bezeichnet den Luftraum um den Flughafen.

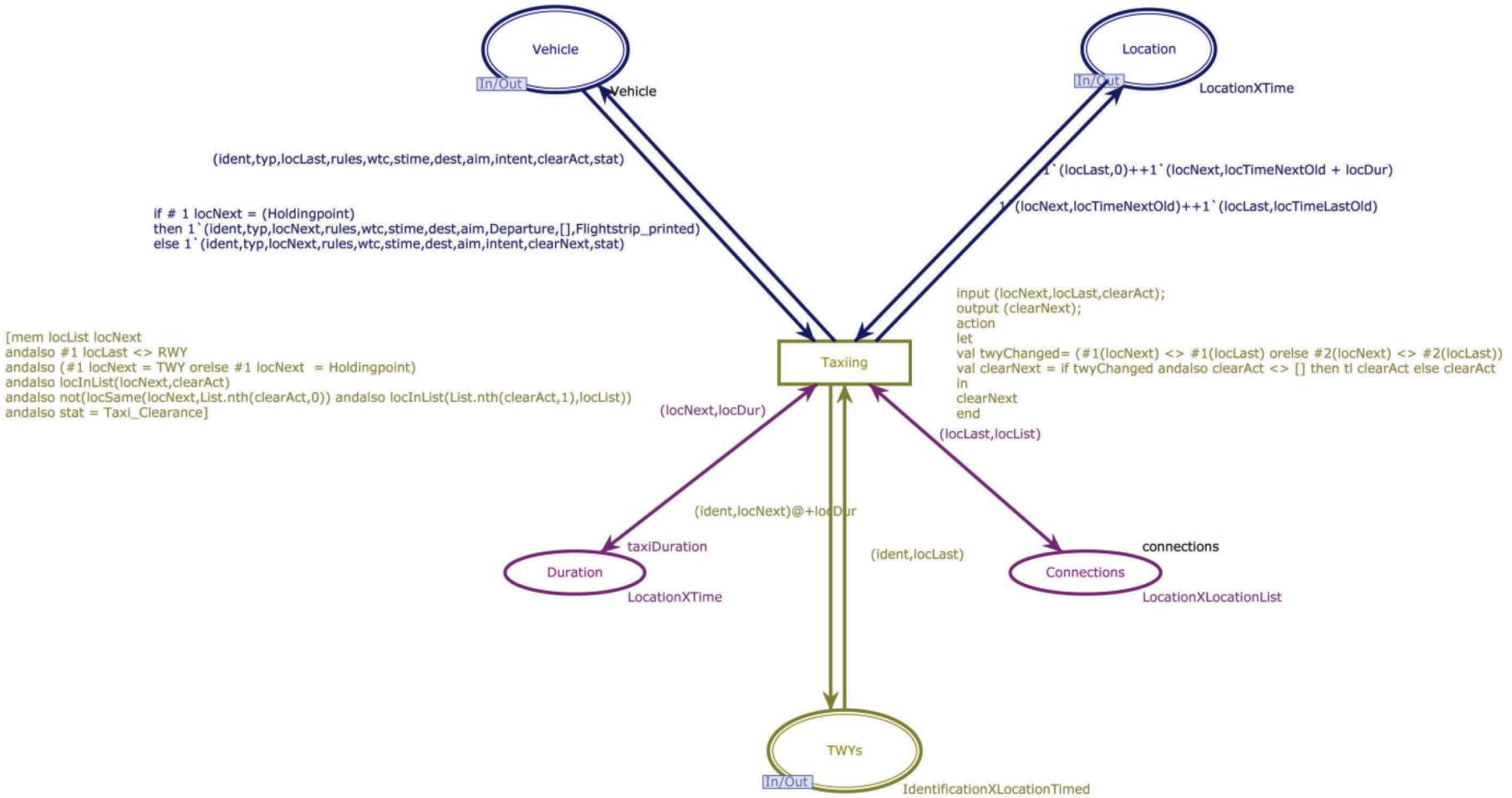

Abbildung 23: Umsetzung der Substitutionstransition Taxiing (Rollen) im Petrinetz

allerdings darauf geachtet werden, dass diese Rollzeiten logisch festgelegt werden. Das bedeutet, dass die tatsächlichen Rollgeschwindigkeiten der Objekte auf einem Flughafen durch das Setzen der Duration auf dem gleichnamigen Platz möglichst genau approximiert werden sollte[34]. Als dritten Teil der variablen Prozessstruktur enthält der Platz `Connections` Informationen darüber, welche Teile des Flugplatzes wie miteinander verbunden sind (vgl. Abbildung 24). Erst hierdurch wird die Struktur des Flughafens hinreichend definiert (Manske, 2013).

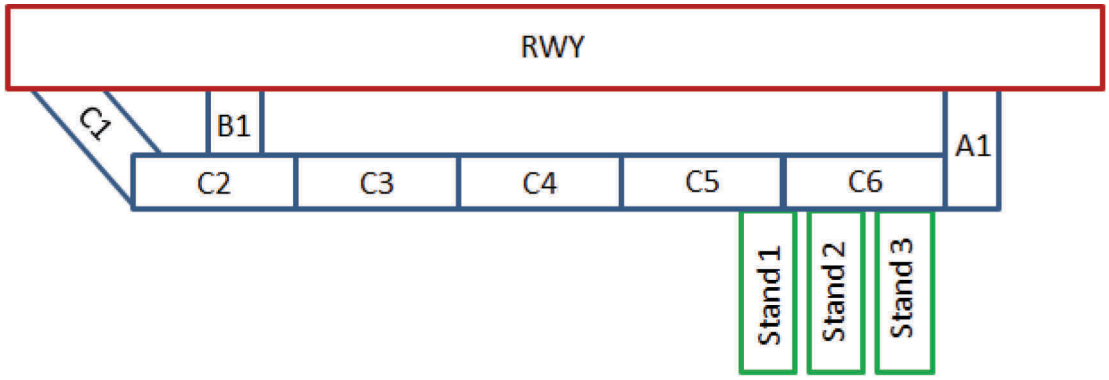

Abbildung 24: Verbindungen zwischen den einzelnen Teilen eines exemplarischen Flugplatzes (eigene Darstellung nach Manske, 2013).

Damit die Bewegung von Objekten über den Flughafen gewährleistet wird, sind die dynamischen Objekte, die sich innerhalb eines bestimmten Szenarios auf dem Flughafen bewegen sollen, auf dem Platz `Vehicles` abgelegt. Hierbei sind alle Informationen der Objekte repräsentiert, die sich in ähnlicher Weise auf dem Flugstreifen eines Lotsen finden lassen. Zusätzlich sind die jeweilige Position, die direkte Intention[35] und die jeweilige Freigabe enthalten. Dies ist im nächsten Abschnitt genauer beschrieben.

Das Rollen eines Objektes erfolgt nun durch das Feuern der Transition `Taxiing` entsprechend der Freigabe und der zuvor definierten Rolldauer. Hierbei wird geprüft, ob eine Freigabe für den gegebenen Rollwegsabschnitt (im Colorset `Vehicle`) enthalten sowie, ob der nächste zu benutzende Rollwegsabschnitt bereits frei ist (Vorhandensein der Marke, die den Rollwegsabschnitt repräsentiert auf dem Platz `Location`). Feuert die Transition, werden die Ortsangaben im Colorset `Vehicle` hinsichtlich des neuen Ortes aktualisiert, sowie die Zeit, zu der die Marke wieder zur Verfügung steht, in Abhängigkeit der definierten Rolldauer heraufgesetzt. Zusätzlich wird der Rollwegsabschnitt, der verwendet wird, als belegt markiert (durch Hochsetzen der Locationzeit).

[34] Diese Approximation kann gemeinhin anhand der vorgeschriebenen Rollgeschwindigkeit von 20 Knoten auf dem Vorfeld, bis maximal 30 Knoten auf dem Rollfeld (Huber, 2012) bzw. 10 Knoten in Kurven (Ferdinand Behrendt, persönliches Gespräch am 27.03.2013) sowie unter Berücksichtigung der jeweiligen Länge des Rollwegsabschnittes erfolgen.

[35] Die Intention beschreibt das Vorhaben eines Luftfahrzeuges, also beispielsweise zu landen.

5.2.3 COLORSETS UND EINLESEN DER INITIALEN MARKIERUNG

Zusätzlich zu den Standard-Colorsets, die CPN-Tools bereitstellt, werden eine Vielzahl weiterer luftfahrtspezifischer Colorsets definiert (Manske, 2013), mit denen auch das Lotsenmodell arbeitet. Die Wichtigsten sollen im Folgenden kurz erläutert werden. Diese können in der Modelldatei im Programmierbereich unter `Declarations` → `Airport` → `Colsets` gefunden werden.

Jeder Flughafenabschnitt (`colset Location`) ist zunächst durch vier Informationen charakterisiert: Typ, Bezeichnung, Nummer und Betriebsrichtung. Dieses setzt sich aus dem Produkt der Colorsets `LocationType`, `LocationName`, `LocationNumber` und `Direction` zusammen.

`LocationType` beschreibt hierbei die Funktion eines Elements z.B. Rollweg oder Stellplatz. Die wichtigsten sind hierbei:

- `TWY`: Rollweg
- `Stand`: Stellplatz
- `RWY`: Piste / Start- und Landebahn
- `Final`: Letzer Teil des Anfluges
- `Crosswind`: Querabflug bei einer Platzrunde[36]
- `Downwind`: Gegenanflug bei einer Platzrunde
- `Base`: Queranflug bei einer Platzrunde
- `Departure`: Abflug / nach dem Start
- `Sky`: Ein LFZ hat den Luftraum der Kontrollzone verlassen, bzw. befindet sich in weiterer Umgebung um den Flughafen.

`LocationName` weist jedem Abschnitt einen Namen zu, anhand dessen verschiedene Elemente des gleichen Typs unterschieden werden können z.B. TWY, A für den Rollweg A. Für die Start-/Landebahnen wird hauptsächlich die Lage verwendet (links L, rechts R), der Name kann hier aber prinzipiell frei gewählt werden.

Durch die Unterteilung der einzelnen Rollwege in kleinere Abschnitte, wird jedem Abschnitt eine Nummer zugeordnet (`LocationNumber`). Diese Nummerierung kann selbst festgelegt werden, sollte aber sinnvoll sein. Die Verbindungen zwischen den einzelnen Abschnitten werden aber nicht anhand der Nummern festgelegt, sondern anhand der `Connections`. Im Fall der Start-/Landebahnen wird die Nummer anhand der Pistenausrichtung vergeben (am Platz EDVE entspricht Ausrichtung 26 einer westlichen Richtung von 260°; Manske, 2013).

Über das Colorset `Direction` wird die Betriebsrichtung der Pisten eines Flughafens festgelegt. Die Betriebsrichtung `Std` wird anhand der primären Betriebsrichtung festgelegt,

[36] Platzrunde bezeichnet ein standardisiertes An- und Abflugverfahren von Flügen nach Sichtflugregeln (VFR) und dient u. A. der Einleitung eines sicheren Landeanfluges.

welche sich nach der häufig vorherrschenden Windrichtung richtet. Am Platz EDVE ist dies die Ausrichtung 26, in die meistens gestartet und gelandet wird. Hierdurch erhalten die Rollwege ebenfalls eine primäre Betriebsrichtung. Im Petrinetz wird zwischen `Std` (standard), `Rev` (reverse/entgegengesetzt) und `Both` (beide Richtungen) unterschieden. Diese Unterscheidung ist für die Verbindung der einzelnen Rollwegsabschnitte wichtig. Zwei nebeneinander liegende Rollwege können in zwei Richtungen befahren werden, wodurch sich andere Verbindungen ergeben.

Zur Bestimmung der dynamischen Objekte, die sich auf dem Flughafen bewegen, wird das Colorset `Vehicle` definiert. Dies enthält alle Informationen zu einem Luftfahrzeug, ähnlich einem Flugstreifen und wie sie auch durch die Analyse von Human-Factors-Consult (2009) beschrieben wurden:

- `Identification`: Rufzeichen
- `Type`: Flugzeugtyp bzw. Fahrzeug z.B. A319; Car
- `Location`: Aktuelle Position des Luftfahrzeuges
- `Rules`: Flugregeln (IFR oder VFR)
- `WTC`: Wirbelschleppenkategorie (Light, Medium, Heavy, Super)
- `S_Time`: Geplante Abflug- oder Ankunftszeit
- `Location`: Zielort; bei abfliegenden Flugzeugen zunächst der Haltepunkt der Piste, anschließend „Sky“; bei ankommenden Flugzeugen die Parkposition
- `Aim`: Ziel; primär An- und Abflug (Arrival; Departure) aber auch Auftanken (Refuel) oder Check (Überprüfen der Piste; für Wartungsfahrzeuge) möglich
- `Intention`: Absichten, die das Fahrzeug verfolgt, z.B. Rollen zur Startbahn, Start, Landung. Diese teilt das Flugzeug dem Lotsen mit, der daraufhin seine Freigabe erteilt.
- `LocationList`: Freigegebene Flughafenabschnitte; enthält eine Liste von Flughafenabschnitten, die das Luftfahrzeug nach Freigabe des Lotsen verwenden darf bzw. in die es einfliegen darf.
- `Status`: Gibt an, welche Art Freigabe ein Luftfahrzeug besitzt z.B. zum Rollen oder Starten

Das Einlesen der initialen Markierungen aller genannten Plätze des Flughafenmodells erfolgt durch die Einbindung der externen Textdateien `Scenario.txt` sowie `Layout.txt`[37]. Durch

[37] Beispiele für diese Dateien befinden sich im Anhang.

die Verwendung sogenannter Structured Colorsets werden die über ein Flugzeug zur Verfügung stehenden Informationen in das Colorset `vehicle` eingebunden, wobei z.B. genau beschrieben werden kann, wo sich welches Luftfahrzeug auf dem Flughafen befindet, welche Aktionen dieses ausführen möchte und welche Freigaben es bereits besitzt.

Mit dem Befehl `val vehicles` wird hierbei die Konstante `vehicles` definiert und festgelegt, welche Marken (also welche konkreten Flugzeuge) diese enthalten soll. Hierbei werden alle Informationen über das Szenario gegeben, also z.B. welche Flugzeuge auftreten sollen, wo diese sich bei Szenario-Beginn befinden, wann diese ihre Anfrage stellen, usw. Weiterhin werden auf die gleiche Weise die Konstanten `twys`, `stands`, `rwys`, `airborne` und `final` bestimmt, welche die initiale Markenbelegung der gleichnamigen Plätze der invariablen Prozesslogik bestimmen.

Die `Layout.txt` enthält alle Informationen über den Aufbau des Flughafens. Vor allem wird hier definiert, welche Rollwege der Flughafen besitzt und wie diese in Abschnitte unterteilt sind (`val location`), mit welchen anderen Rollwegen der betreffende Abschnitt jeweils verbunden ist (`val connections`) und wie lange das Überqueren des einzelnen Rollwegsabschnittes dauert (`val taxiDuration`). Bei der Erstellung eines Flughafenmodells ist daher im Vorfeld anhand der Flughafendiagramme genau festzulegen, welche Locations definiert werden sollen, wie diese in Abschnitte unterteilt werden und wie diese mit anderen Locations verbunden sind. Ferner ist zu berechnen, wie lange ein Luftfahrzeug jeweils benötigt, um einen bestimmten Rollwegsabschnitt zu überqueren/ zu durchfahren[38].

5.3 Das Interaktionsmodell

5.3.1 Theoretische Erläuterung

Das Interaktionsmodell dient dem Austausch von Informationen zwischen dem Lotsen und dem Flughafen. Hierbei wurden die aus der Feldstudie der Arbeitsumgebung wichtigsten Informationsquellen modelliert: Die Außensicht (bei schlechter Sicht auch Radar[39]), die (neu erhaltenen) Flugstreifen, sowie der Funkkanal. Eine Interaktion kann hierbei von beiden Seiten initiiert werden. Der Prozess kann dem Lotsen Informationen zur Verfügung stellen, wie auch der Lotse dem Prozess Informationen geben kann. Hierbei kann der Lotse vom Prozess Informationen geliefert bekommen (Information Push; e.g. Cybenko & Brewington, 1999), kann aber auch aktiv Informationen anfordern (Information Pull). Der Informationsfluss vom Prozess zum Lotsen (Push) erfolgt einerseits durch das Drucken neuer Flugstreifen und andererseits

[38] In der Version 1.9 wurde die Zeit anhand der zulässigen Rollwegsgeschwindigkeit von 20 Knoten auf geraden und 10 Knoten auf Kurvenabschnitten (Ferdinand Behrendt, persönliches Gespräch am 27.03.2013) und einer durchschnittlichen Länge der Rollwegsabschnitte von 100 m approximiert.
[39] Das Radar wurde in der vorliegenden Modellversion lediglich als Ersatz der Außensicht bei schlechter Sicht eingefügt. Es werden bei beiden redundante Informationen zur Verfügung gestellt.

durch das Stellen von Requests über den Funkkanal. Beide führen zu einer gezielten Reaktion des Lotsenmodells (siehe Abschnitt 5.4). Zusätzlich kann und muss das Lotsenmodell gezielt Informationen aus dem Interaktionsmodell abrufen, um seine Aufgabe erfüllen zu können (Pull). So müssen über die Außensicht (bzw. bei schlechter Sicht das Radar) Informationen über die Zustände bestimmter Locations eingeholt werden, damit das Lotsenmodell entscheiden kann, ob eine Freigabe gegeben werden kann oder nicht (vgl. Abschnitt 5.4.6).

Der Funkkanal wird außerdem vom Lotsenmodell verwendet, um seinerseits Informationen an den Prozess zu senden. Konkret bedeutet dies, dass vom Lotsen Freigaben an das jeweilige Luftfahrzeug übermittelt werden. Neben dem Komplexitätsfaktor Verkehrsaufkommen, der über das Flughafenmodell beeinflusst werden kann, werden hier zwei weitere Komplexitätsfaktoren implementiert: Die Sichtbedingungen (die in der vorliegenden Untersuchung keine Rolle spielen), sowie die Belastung der Funkfrequenz, die einerseits vom Verkehrsaufkommen und andererseits der Länge der jeweiligen Kommunikation abhängig ist (vgl. Kapitel 4.2).

5.3.2 Das Petrinetz-Modell im Detail

Im Zentrum des Interaktionsmodells steht der Platz `Vehicles` (vgl. Abbildung 25 auf Seite 97). Hiervon ausgehend erfolgen alle Aktionen, die aus dem Flughafenmodell heraus angestoßen werden. Hierbei handelt es sich um das Drucken der Flugstreifen über die Transition `print_Flightstrip`, sowie das Stellen von Anfragen über die Transition `Requesting_TWR`. Hierbei ist definiert, dass die Transition `print_Flightstrip` erst dann feuern kann, wenn die Zeit @x[40] einer Marke auf dem Platz `Vehicles` den Wert der Modellzeit erreicht hat (das Token steht zur Verfügung; vgl. Kapitel 4.6). Ferner muss die Marke, die ein Flugzeug repräsentiert den Status `Airborne`, `Waiting` oder `Approaching` enthalten. Ist dies gegeben, so wird der Druck eines Flugstreifens ausgelöst. Durch die direkte Rückgabe der Marke auf den Platz `Vehicles` wird der Status der Marke, die das jeweilige Flugzeug repräsentiert, auf `Flightstrip_printed` gesetzt, so dass verhindert wird, dass ein weiterer Flugstreifen für dasselbe Flugzeug gedruckt wird. Ferner bietet dies die Möglichkeit, dem Flugzeug eine Zeitverzögerung hinzuzufügen[41].

Die Transition `Requesting_TWR` kann nun als nächstes auftreten, sofern folgende Bedingungen erfüllt sind: Eine eventuelle zusätzliche Verzögerungszeit ist verstrichen (@x ≤ Modellzeit), der Status wurde auf `Flightstrip_printed` aktualisiert und es ist eine Marke

[40] x Bezeichnet hierbei eine beliebige Zahl.

[41] Um zu simulieren, dass der Flugstreifen etwa fünf Minuten vor dem eigentlichen Eintreffen eines Flugzeuges gedruckt wird, wurde der Transition `print_Flightstrip` in früheren Modellversionen eine Zeitverzögerung von @+300 (+300 Sekunden; vgl. Kapitel 4.6) hinzugefügt. In Version 1.9, die zur Modellvalidierung diente, entfällt diese Verzögerung, da die Versuchspersonen die Flugstreifen bereits vor sich liegen haben. Das Modell erhält den Flugstreifen unverzüglich bei Eintreffen des jeweiligen Flugzeuges.

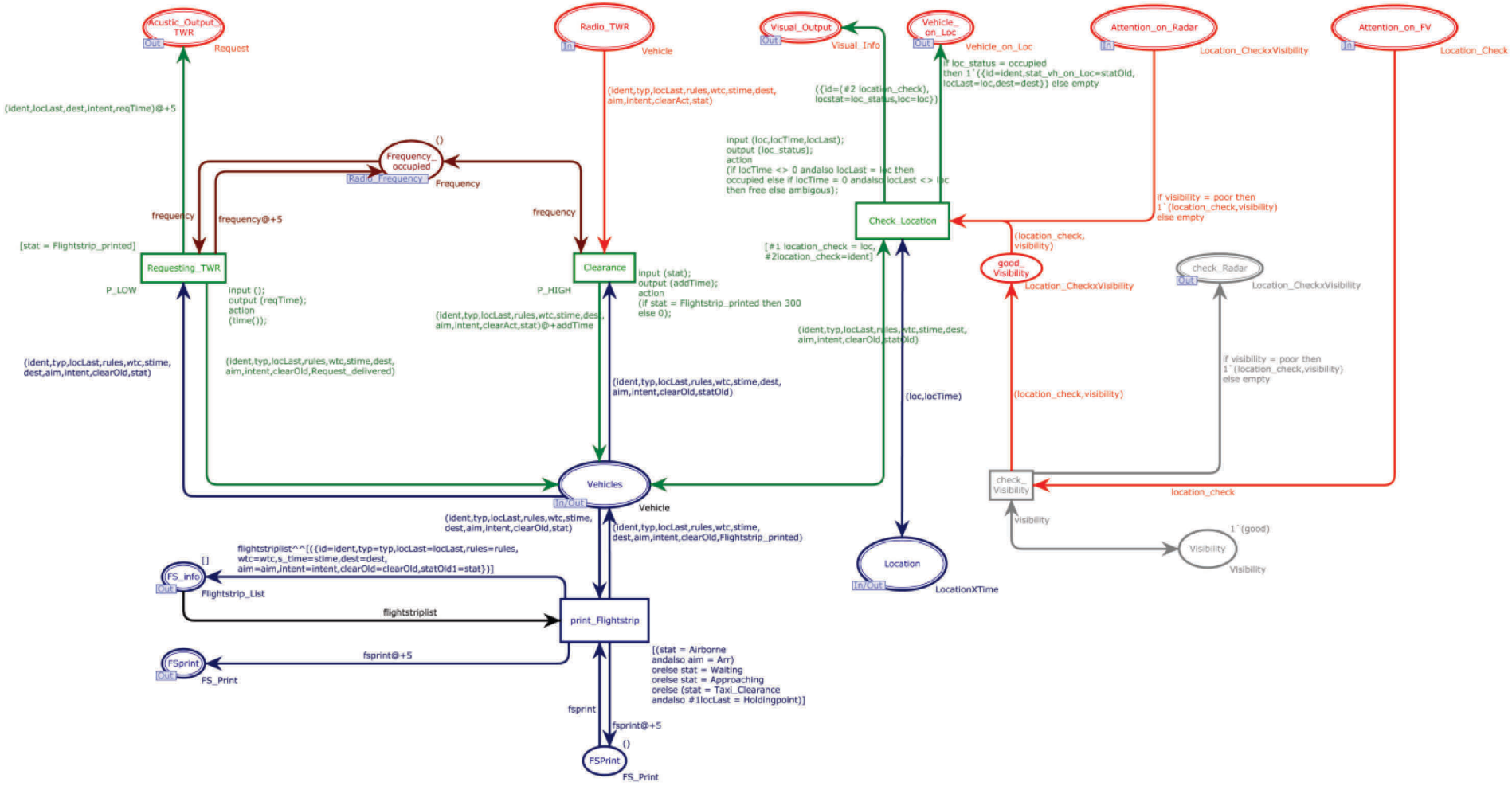

Abbildung 25: Das Interaction-Modell im Petrinetz

auf dem Platz Frequency_occupied vorhanden, dessen Zeit gleich der Modellzeit ist (@x ≤ Modellzeit). Letzteres beschreibt die serielle Verwendung des Funkkanales. Ist dies gegeben, wird damit eine Marke auf dem Platz Acoustic_Output_TWR generiert, welche die Anfrage eines Flugzeuges symbolisiert. Darüber hinaus wird die Zeit der Marke auf dem Platz Frequency_occupied erhöht. Die Länge dieser Zeit entspricht damit der Länge einer einzelnen Anfrage bzw. Anweisung und wird konstant mit einer Länge von fünf Sekunden modelliert, was in etwa der mittleren Länge aller Anfragen und Anweisungslängen des Validierungsversuches entspricht.

Nach der Verarbeitung aller nötigen Informationen und dem Treffen einer Entscheidung durch das Lotsenmodell erscheint auf dem Platz Radio_TWR eine Marke, welche das Vorhandensein einer Freigabe symbolisiert. Ist zusätzlich die Zeit der Marke auf dem Platz Frequency_occupied gleich oder kleiner der Modellzeit, so kann die Transition Clearance feuern und die Freigabe wird an das Flughafenmodell (das Flugzeug) übermittelt. Hierbei werden alle nötigen Einträge in der jeweiligen Marke auf dem Platz Vehicles aktualisiert (also beispielsweise, um welche Freigabe es sich handelt und für welche Locations die Freigabe gilt). Für das Treffen von Entscheidungen muss das Lotsenmodell bestimmte Informationen sammeln. Hierzu gehören Informationen über den Zustand bestimmter Locations des Flughafenmodells beispielsweise, ob die Start-/Landebahn frei oder belegt ist. Mit einem Blick aus dem modellierten Fenster kann das Lotsenmodell diese Information aufnehmen.

Liegt die visuelle Aufmerksamkeit auf der Außensicht, erscheint eine Marke auf dem Platz Attention_on_FV (FV für far view – Außensicht). Bei guten Sichtbedingungen, die über den Platz Visibility eingestellt werden können, kann ein direkter Abruf der Zustände der benötigten Location durch feuern der Transition Check_Location vom Platz Location erfolgen. Für den Fall, dass die Location (z.B. die Start-/Landebahn) frei ist, wird eine Marke auf dem Platz Visual_Output generiert, welcher die Informationen enthält, welchen Status die jeweilige Location besitzt (also z.B. RWY, free, wenn die Piste frei ist). Ist die Piste belegt, so wird zusätzlich zu der Marke auf dem Platz Visual_Output eine weitere Marke auf dem Platz Vehicle_on_Loc erstellt, um zu symbolisieren, welches Luftfahrzeug sich auf der jeweiligen Location befindet. Gleichzeitig wird in die Marke auf Visual_Output geschrieben, dass die Location belegt ist (also z.B. RWY, occpied). Je nachdem, welcher Satus hierbei generiert und an das Lotsenmodell weitergeleitet wird, werden unterschiedliche Handlungen vom Lotsenmodell erforderlich. Dies wird im folgenden Abschnitt näher erläutert.

5.4 DAS LOTSENMODELL

5.4.1 THEORETISCHE ERLÄUTERUNGEN

Das Lotsenmodell stellt die eigentliche kognitive Simulation dar und bildet die kognitiven Prozesse des Lotsen ab. Zusätzlich werden aber auch kognitive Ressourcen modelliert, welche für den vorliegenden Anwendungsfall speziell von Interesse sind.

Grundlegend orientiert sich das Lotsenmodell an der Handlungsregulationstheorie (Hacker, 1986), sowie an der in Kapitel 4.3 durchgeführten Aufgabenanalyse und dem in Kapitel 4.5 beschriebenen theoretischen Modell. Dabei wurde das Lotsenmodell nicht exakt in dem Detailgrad und der Tiefe umgesetzt, wie in der Aufgabenanalyse und dem theoretischen Modell beschrieben, sondern nur so detailliert wie nötig.

Die einzelnen Phasen der Handlungsregulation (wie in Kapitel 4.5 beschrieben) wurden im Modell als Sub-Module umgesetzt. Als zusätzliches Sub-Modul wurde ein Arbeitsgedächtnis-Modul (`working_memory`) abgebildet, welches die Speicherung und den Abruf von Informationseinheiten beschreibt und eine begrenzte Kapazität besitzt.

In Anlehnung an die Handlungsregulationstheorie nach Hacker (1986) wurden die Phasen Wahrnehmung (`perceive`; Abschnitt 5.4.2 und 5.4.3), Planung (`plan`; Abschnitt 5.4.4 und 5.4.5), Entscheidung (`decide`; Abschnitt 5.4.6 und 5.4.7) und Handlungsausführung (`act`; Abschnitt 5.4.8 und 5.4.9) als Sub-Module realisiert (vgl. Abbildung 26 auf Seite 100), nach denen auch die einzelnen Unteraufgaben und kognitiven Prozesse im theoretischen Modell systematisiert worden sind. Wie im theoretischen Modell beschrieben, dient die Phase der Wahrnehmung der Aufnahme aller relevanten Informationen, die Phase der Planung beschreibt die Einordnung der Flüge in die Stripbay anhand ihrer Ankunfts bzw. Abflugszeit. In der Entscheidungsphase wird auf Grundlage der aufgenommen Informationen und der Planung festgelegt, welchem Flugzeug als nächstes welche Freigabe erteilt wird. Diese wird in der Phase der Handlungsausführung übermittelt.

Da die Phasen, wie bereits beschrieben und aus der Aufgabenanalyse ersichtlich, nicht streng sequentiell ablaufen, sind Sprünge zwischen den einzelnen Phasen möglich. So müssen z.B. auch für die Phase der Planung Handlungen ausgeführt werden; zur Entscheidung sind Informationen nötig, die erst aktiv aufgenommen werden, wenn die jeweilige Entscheidung getroffen werden muss. Im Folgenden werden die einzelnen Phasen im Detail vorgestellt.

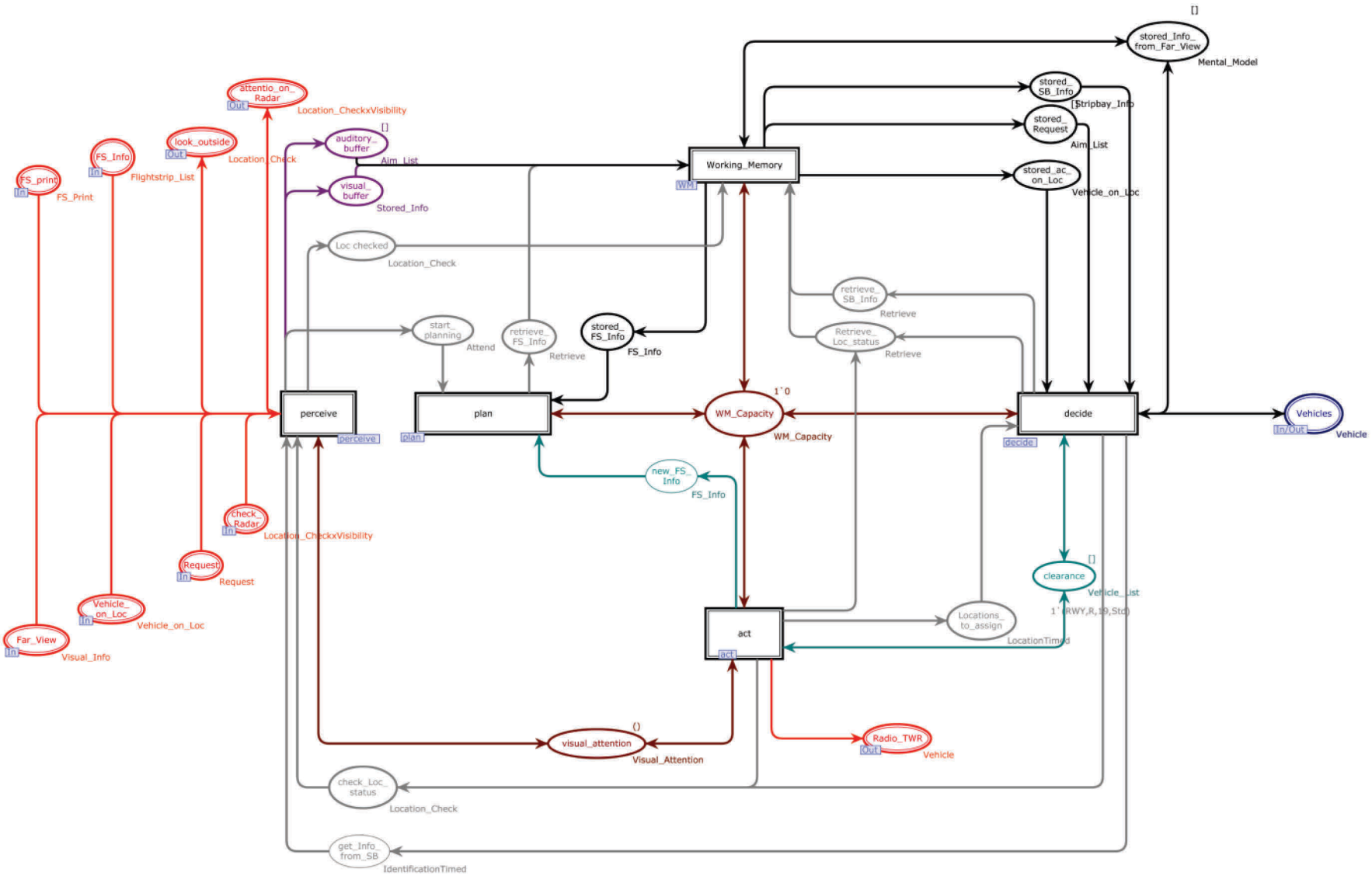

Abbildung 26: Das Lotsenmodell im Petrinetz

5.4.2 PHASE DER INFORMATIONSAUFNAHME

Die Phase der Informationsaufnahme (`perceive`) dient der Aufnahme aller relevanten Informationen, wie sie durch die Aufgabenanalysen beschrieben werden. Die aktive Aufnahme von Informationen wird hierbei durch das Eintreffen von Informationen aus dem zu kontrollierenden Prozess getriggert. Das Sub-Modul `perceive` ist seinerseits nochmals in zwei weitere Sub-Module unterteilt, welche die Aufnahme von Informationen mittels verteilter (`devided`) und fokussierter (`focused`) Aufmerksamkeit darstellen **(Annahme 1)**. Diese Unterteilung orientiert sich an der Annahme von Wickens (2002), dass neben den verschiedenen Modalitäten (auditiv und visuell) zwei Aspekte der visuellen Verarbeitung, focal und ambient, existieren. Visuelle Stimuli können in beiden Aspekten aufgenommen werden (verteilt und fokussiert), wobei sich der Informationsgehalt der aufnehmbaren Stimuli unterscheidet. Fokussierte Aufmerksamkeit wird benötigt, um Details und Muster zu erkennen, verteilte Aufmerksamkeit hingegen dient der Hinwendung zu Objekten (Wickens, 2002; **Annahme 2**). Auditive Stimuli werden hier als in nur einem Aspekt aufnehmbar modelliert, welcher hier ebenfalls als verteilte Aufmerksamkeit bezeichnet wird, einerseits da der Hörsinn ein Fernsinn ist. Andererseits kann davon ausgegangen werden, dass der Lotse aufgrund der Wichtigkeit des Funkkanales permanent seine auditive Aufmerksamkeit willentlich auf den Funkkanal gerichtet hat (vgl. Broadbent, 1954, 1958; Treisman, 1960; **Annahme 3**).

Im Sub-Modul `devided` wurden daher zwei Stimuli modelliert, welche mit verteilter Aufmerksamkeit aufgenommen werden können: Diese sind 1) die Information, dass ein Flugstreifen gedruckt wurde über beide Kanäle (auditiv und visuell) und 2) die Anfrage (request) eines Piloten über den auditiven Kanal. Die *Information über den Druck eines Flugstreifens* ist hierbei eine sehr einfache Information, welche vornehmlich dazu dient, eine visuelle Aufmerksamkeitsverschiebung zu initiieren, also die Hinwendung zu einem Objekt zu triggern, weshalb sie mit verteilter Aufmerksamkeit wahrgenommen werden kann (vgl. Wickens, 2002). Die Information darüber, dass ein Flugstreifen gedruckt wurde, kann hierbei sowohl auditiv als auch visuell wahrgenommen werden, sodass bei Belegung eines Kanals der jeweils andere die Information dennoch extrahieren kann. Das Bemerken eines Flugstreifens, der gedruckt wird, stellt hierbei einen Bottom-Up[42]-Trigger zur Aufmerksamkeitsverschiebung dar. Um den vollen Informationsgehalt des gedruckten Flugstreifens, also die detaillierten darauf enthaltenen Informationen zu erfassen, ist eine Verschiebung der visuellen Aufmerksamkeit auf den Flugstreifen notwendig (Fokussierung). Wird diese auf den Flugstreifen gerichtet, muss er zunächst aus dem Drucker entnommen werden (vgl. auch Kapitel 5.4.8. und 5.4.9), bevor die

[42] Als Bottom-Up werden Handlungen bezeichnet, welche hauptsächlich durch das Auftreten von externen Hinweisreizen und eingehenden Informationen initiiert werden (Kallus et al., 1999).

Informationen extrahiert werden können und sie für den Planungsprozess zur Verfügung stehen. Wie der Prozess im Detail modelliert wurde wird im folgenden Abschnitt 5.4.3 erläutert.

Die *Wahrnehmung der Anfrage eines Piloten* über den Funkkanal wird über den auditiven Kanal erfüllt. Wie bereits oben beschrieben, wird hier nicht von einer Unterscheidung zwischen fokussierter und verteilter Aufmerksamkeit ausgegangen, sodass die Informationen, die über diesen Kanal kommen bereits vollständig extrahiert werden, ohne dass eine aktive Aufmerksamkeitsverschiebung notwendig ist. Dies stellt hierbei bereits die Aufnahme des kompletten Chunks dar. Unter der Voraussetzung, dass das SEEV-Modell[43] (Wickens & McCarley, 2008; Wickens et al., 2007; Wickens, Helleberg, Goh, Xu, & Horrey, 2001) auch für auditive Reize gilt, kann man davon ausgehen, dass für diese eine vergleichsweise hohe Wahrscheinlichkeit besteht, dass ihnen Aufmerksamkeit geschenkt wird. Dies ergibt sich dadurch, dass die Salienz (Salience) sehr hoch ist, während der Aufwand einer Aufmerksamkeitsverschiebung (Effort) sehr gering ist. Weiterhin ist die Erwartung, auf diesem Kanal wichtige Informationen zu erhalten (Expectancy), sowie auch dessen Wert (Value) sehr hoch. Zusätzlich ergibt sich aus der Beobachtung, dass der Lotse permanent seinen Kommunikationskanal mit dem Piloten überwacht und dieser eine der wichtigsten Informationsquellen des Lotsen darstellt.

Aufmerksamkeitsverschiebungen, wie sie eben beschrieben wurden, können allerdings auch Top-Down[44] initiiert werden, wenn bestimmte Informationen für einen Entscheidungsprozess benötigt werden. So wurde festgelegt, dass für den Entscheidungsprozess drei Informationen vonnöten sind **(Annahme 4)**: Die Anfrage eines Piloten, der korrespondierende Flugstreifen aus der Stripbay, sowie die Information über den Zustand der jeweiligen zu verwendenden `Location` durch einen Blick nach draußen. Hierbei werden für den Abruf der Informationen aus dem Flugstreifen und der Außensicht zwei Top-Down Trigger geschaltet, welche die Aufmerksamkeitsverschiebung zur Stripbay und zur Außensicht initiieren. Die Außensicht hat hierbei Priorität. Durch die ablaufenden Prozesse werden die benötigten Informationen aus der Außensicht und der Stripbay zunächst im Arbeitsgedächtnis gespeichert und anschließend dem nachfolgenden Entscheidungsprozess zur Verfügung gestellt (mehr hierzu im Abschnitt 5.4.7).

5.4.3 Das perceive-Modell im Petrinetz

Das Perceive-Modell besteht aus 21 Plätzen und fünf Transitionen. Die Aufnahme der Informationen aus dem Interaktionsmodell erfolgt zunächst über zwei Plätze im `devided` Sub-

[43] Mit dem SEEV-Modell beschreiben Wickens und Kollegen, wie Menschen ihre (visuelle) Aufmerksamkeit auf verschiedene handlungsrelevante Reize verteilen. Hierbei kann in Abhängigkeit von vier Haupteinflussfaktoren (Salience, Effort, Expectancy, Value) bestimmt werden, wie hoch die Wahrscheinlichkeit ist, dass einem bestimmten Reiz Aufmerksamkeit zugewendet wird: P(A) = sS – efEF + (exEX + vV) (Wickens & McCarley, 2008, S. 56)

[44] Als Top-Down werden Prozesse bezeichnet, welche vornehmlich auf Plänen, Intentionen und Regeln beruhen. Es wird damit Verhalten beschrieben, was hauptsächlich durch Kognition entsteht (Kallus et al., 1999).

Modul (siehe Abbildung 27), welche wie bereits beschrieben die Aufmerksamkeitsverschiebungen des Modells steuern und tiggern: die Information über den Druck eines Flugstreifens (`FS_Print`) und eine Anfrage (`Request`) (in Abbildung 27 und Abbildung 28 auf Seite 104 jeweils rot dargestellt). Hierbei wird ein einfaches Modell zugrunde gelegt, welches die Informationen vor der Aufnahme (Plätze `FS_Print` und `Request`), sowie nach ihrer Aufnahme (Plätze `FS_Print_noticed` und `Auditory_Buffer`) beschreibt. Zusätzlich werden durch die Plätze `visual_attention` und `auditory_attention` die jeweiligen Aufmerksamkeitsressourcen modelliert. Diese sind als gezeitete (timed) Marken abgebildet, sodass diese bei Aufnahme von Informationen für eine bestimmte Zeit gebunden werden, bevor neue Informationen aufgenommen werden können.

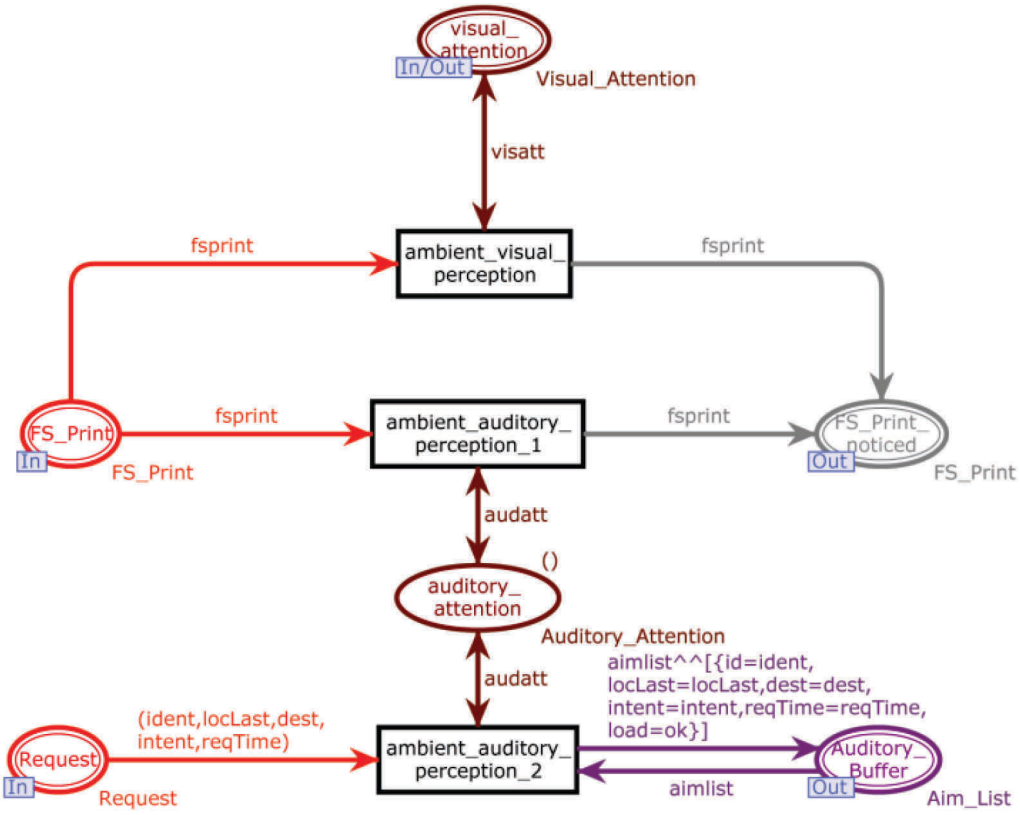

Abbildung 27: Das Devided-Sub-Modul im Petrinetz

Die Transitionen `ambient_visual_perception`, `ambient_auditory_perception_1`, sowie `ambient_ auditory_perception_2` stellen hierbei jeweils den Prozess der eigentlichen Informationsaufnahme dar. Der Einfachheit halber wurden hier zwei Plätze für die verschiedenen Informationen, sowie zwei Transitionen für die auditive Informationsaufnahme modelliert[45].

[45] Eine etwas elegantere Modellierung hätte die Verwendung eines Union Platzes erfordert, wobei hier der Aufwand der Implementierung eines solchen Platzes nicht im Verhältnis zu seinem Nutzen steht.

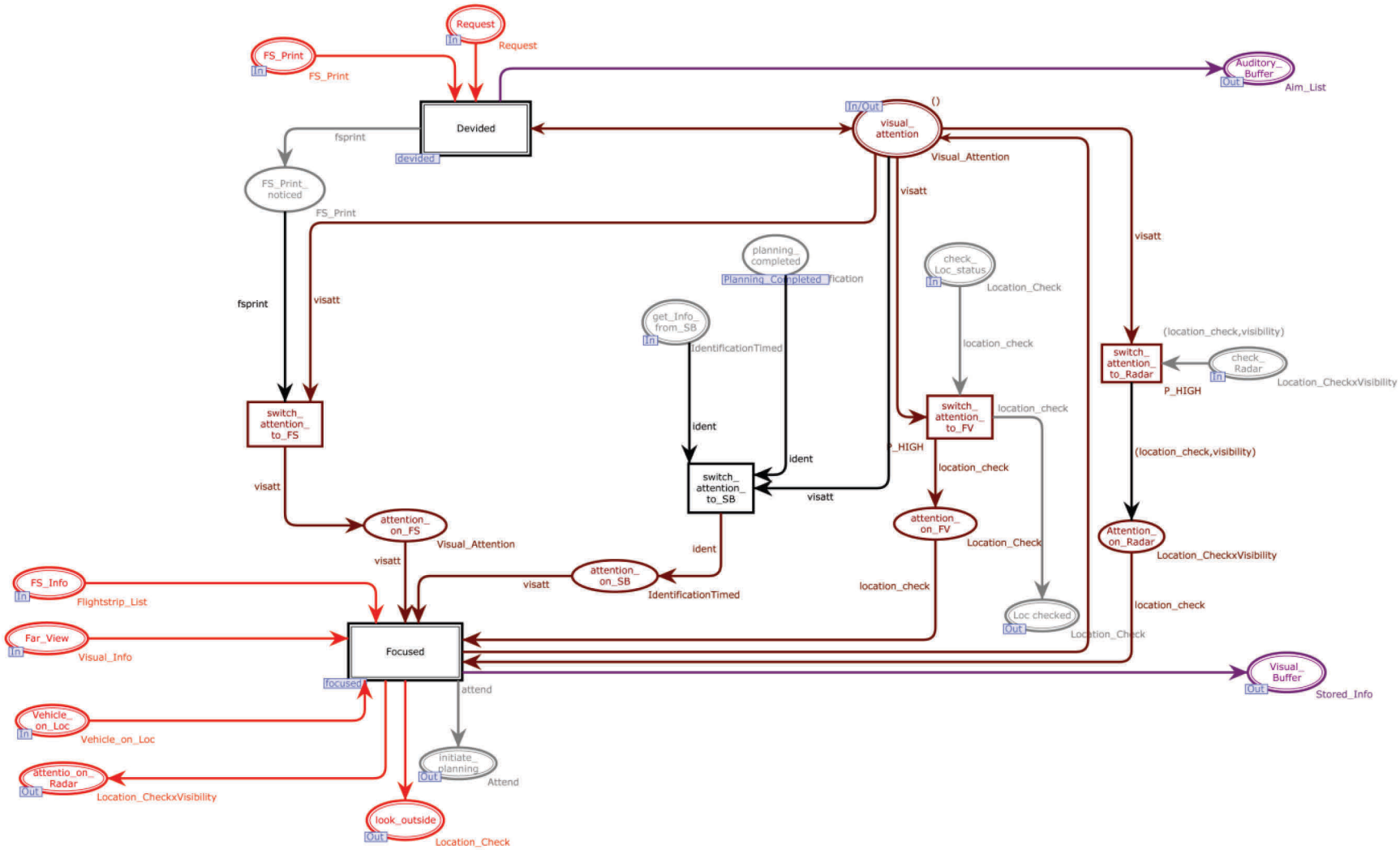

Abbildung 28: Das Perceive-Sub-Modul im Petrinetz

Die so aufgenommenen Informationen werden auf den nachgeschalteten Plätzen abgelegt und stehen dann den nachfolgenden Prozessen zur Verfügung. Diese sind `working_memory` bei `auditory_buffer` (siehe Abschnitt 5.4.10), sowie der Aufmerksamkeitsverschiebung (im Sub-Modul `perceive`) bei `FS_Print_noticed`.

Der Prozess der Aufmerksamkeitsverschiebung zum gedruckten Flugstreifen ist im hierarchisch darüber angeordneten `percieve` Modul modelliert (Abbildung 28 auf Seite 104) und wird durch die Transition `switch_attention_to_FS` abgebildet. Diese Transition konkurriert jeweils um die visuelle Aufmerksamkeit mit drei weiteren Areas of Interest (AOIs): der Stripbay (`switch_attention_to_SB`), der Außensicht (`switch_attention_to_FV`) und dem Radar (`switch_attention_to_Radar`). Diese Aufmerksamkeitsverschiebungen werden, wie bereits erwähnt, jeweils durch Top-Down-Prozesse angestoßen, so dass weitere Bedingungen für deren Auftreten erfüllt sein müssen. So muss für eine Aufmerksamkeitsverschiebung zur Stripbay die Planung abgeschlossen sein, sowie die Informationen aus der Stripbay für eine Entscheidung benötigt werden (der Transition vorgeschaltete Plätze `planning_completed` und `get_Info_from_SB`). Für eine Verschiebung der Aufmerksamkeit zur Außensicht, sowie zum Radar muss ebenfalls die jeweilige Information für eine Entscheidung benötigt werden (vorgeschaltete Plätze `check_Loc_status` und `check_Radar`). Hierbei wird wiederum in Anlehnung an das SEEV-Modell (Wickens & McCarley, 2008; Wickens et al., 2007, 2001) sowie an Untersuchungen von Pinska und Bourgois (2005) sowie Pinska (2008) der Außensicht und dem Radar eine höhere Priorität zugeordnet, dass die Aufmerksamkeit dorthin verschoben wird, sollte die Auswahl zwischen mehreren AOIs bestehen. Dies ergibt sich insbesondere daraus, dass der Informationsgehalt (Value) welcher aus der Außensicht bzw. dem Radar erhalten wird (ein Überblick über die aktuelle Verkehrssituation), sowie die Erwartung, auf diesem Kanal wichtige Informationen zu erhalten (Expectancy), als höher angenommen werden kann, als es bei den anderen beiden AOIs der Fall ist **(Annahme 5)**. Die Salienz und der Aufwand können hierbei als gleich angesehen werden. Die Zuordnung einer solchen Priorität erfolgt qualitativ[46] durch das Setzen des Prioritätswertes der Transitionen `switch_attention_to_FV` und `switch_attention_to_Radar` auf `P_HIGH`, welcher in CPN-Tools voreingestellt ist. Die beiden anderen Transitionen behalten ihren Wert `P_NORMAL` (wird in CPN-Tools nicht extra angezeigt).

[46] Das Setzen der Priorität könnte prinzipiell auch quantitativ erfolgen, wenn man die exakte Wahrscheinlichkeit für eine Aufmerksamkeitsverschiebung z.B. nach dem SEEV-Modell genau berechnet und diese anstatt der Standard Priorities in CPN-Tools übernimmt. Ebenso können die bei Pinska und Bourgois (2005), Pinska und Tijus (2007) sowie Pinska (2008) gegebenen Wahrscheinlichkeiten für die Übergänge zwischen einzelnen Informationsquellen übernommen werden.

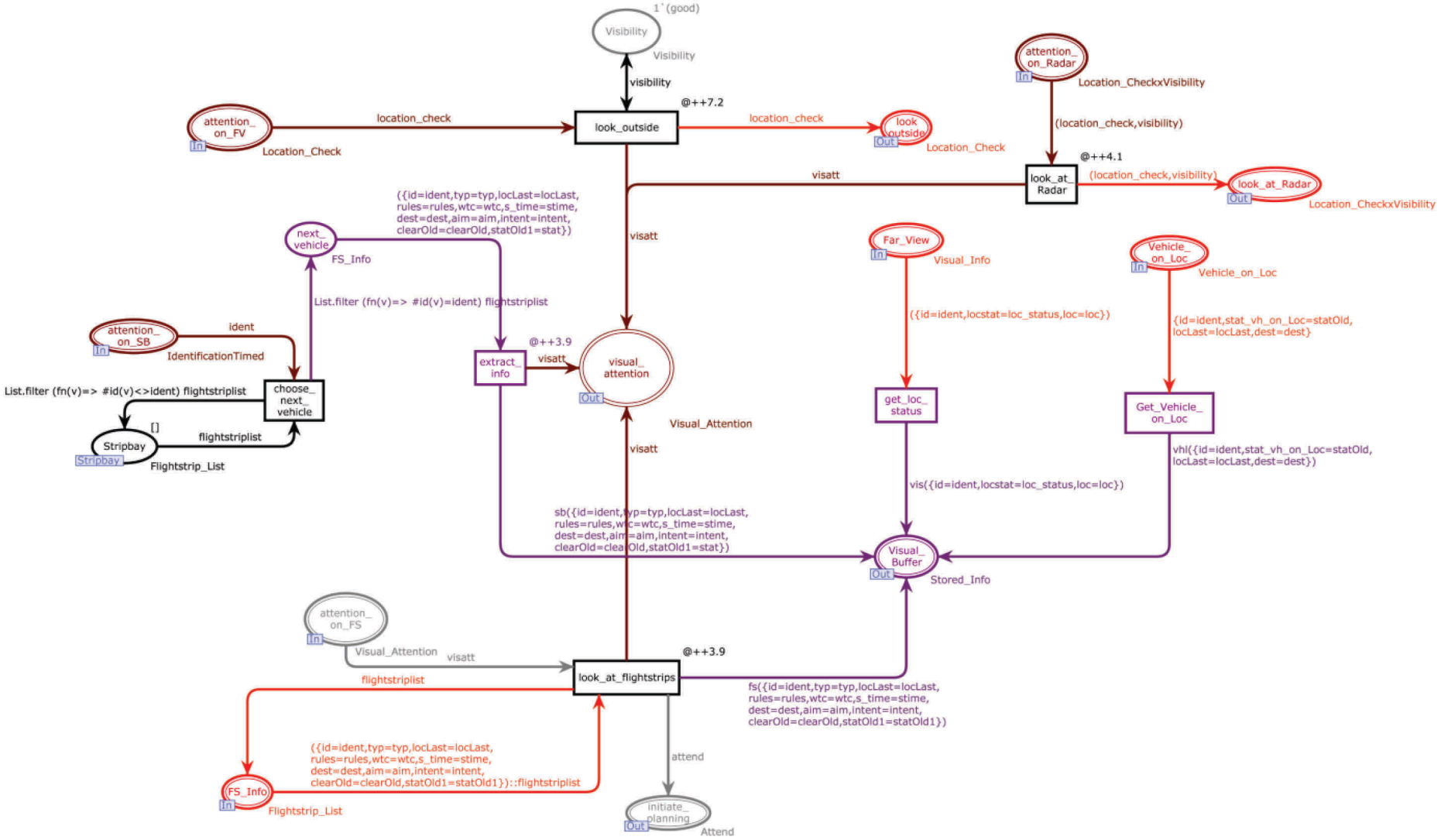

Abbildung 29: Das focused-Sub-Modul im Petrinetz

Im Anschluss an die Aufmerksamkeitsverschiebung können die jeweiligen Prozesse im `focused` Sub-Modul ablaufen (siehe Abbildung 29 auf Seite 106).

Je nachdem, wohin die Aufmerksamkeit im vorangegangenen Schritt verschoben wurde, können dann die spezifischen Informationen extrahiert werden. Durch die Verschiebung zum neu gedruckten Flugstreifen können beispielsweise die genauen darauf enthaltenen Informationen extrahiert werden. Hierbei können immer nur die Informationen von genau einem Flugstreifen in der Liste extrahiert werden, wobei es sich immer um das erste Element der geordneten Liste, die auf dem Platz `FS_Info` liegt, handelt. Bei vorhandenem Flugstreifen führt die Aufmerksamkeitsverschiebung zum neu gedruckten Flugstreifen zur Aktivierung der Transition `look_at_flightstrip`, wodurch einerseits die extrahierte Information in den nachgeschalteten Platz `Visual_Buffer` gelegt wird und andererseits der Trigger zur Initiierung der Verkehrsplanung gesetzt wird (nachgeschalteter Platz `initiate_planning`)[47]. Hierbei wird eine Verzögerung von 3,9 Sekunden hinzuaddiert, was der mittleren Zeit entspricht, die Lotsen in Felduntersuchungen für den Blick auf die Flugstreifen benötigen (Pinska, 2008; **Annahme 6**).

Die Aufmerksamkeitsverschiebungen zu den anderen AOIs erfolgt durch Top-Down-Trigger. Hierbei wird jeweils auf den Plätzen `check_Loc_status`, sowie `get_Info_from_SB` eine Marke generiert. Bei freier visueller Aufmerksamkeit auf dem Platz `visual_Attention`, sowie abgeschlossener Planung (Marke auf dem Platz `planning_completed`) können die Transitionen `switch_attention_to_FV` und `switch_attention_to_SB` feuern, wobei `switch_attention_to_FV` wie bereits beschrieben eine höhere Priorität zugeordnet ist. Ist die Aufmerksamkeit verschoben, so werden im `focused` Sub-Modul Marken auf die Plätze `attention_on_FV` bzw. `attention_on_SB` generiert, wodurch die nachgeschalteten Transitionen `look_outside` bzw. `choose_next_vehicle` feuern können. Bei Auftreten der Transition `look_outside` wird eine Marke auf den nachgeschalteten Plätzen `look_outside` sowie `visual_attention` generiert, wobei beiden eine Verzögerung von 7,2 Sekunden hinzugefügt wird **(Annahme 7)**. Dieser in Felduntersuchungen ermittelte Wert entspricht der mittleren Zeit, die ein Lotse für den Blick aus dem Fenster benötigt (Pinska, 2008). Die visuelle Aufmerksamkeit ist damit für 7,2 Sekunden für den Prozess des „aus-dem-Fenster-Schauens“ gebunden. Durch das Erscheinen einer Marke auf dem Platz `look_outside` wird auf dem Platz `Attention_on_FV` im Interaction-Modell eine Marke abgelegt, so dass der Prozess des Informationsabrufes aus dem Flughafenmodell, wie er in Kapitel 5.3.2 beschrieben ist, erfolgen kann. Sind diese Prozesse ausgeführt, werden je nach Zustand der überprüften `Location` eine

[47] In Version 1.8 ist vor der Aufmerksamkeitsverschiebung zunächst durch vorgeschaltete motorische Aktionen der Flugstreifen aus dem Drucker zu entnehmen und in die Leseposition zu führen.

Marke auf dem Platz `Far_View` und `Vehicle_on_Loc` generiert. Bei freier `Location` wird dabei eine Marke auf dem Platz `Far_View` generiert, welche unter anderem den Status „free“ der `Location` beinhaltet. Bei belegter `Location` wird eine Marke auf dem Platz `Far_View` generiert, welche den Status „occupied“ enthält. Zusätzlich wird auf dem Platz `Vehicle_on_Loc` eine Marke generiert, welche die Information enthält, welches Luftfahrzeug die `Location` gerade besetzt. Die so erhaltenen Informationen werden im Anschluss durch das Feuern der Transitionen `get_loc_status` bzw. `Get_Vehicle_on_Loc` in den `Visual_Buffer` überführt. Bei Auftreten der Transition `choose_next_vehicle` wird der auf dem Platz `Stripbay` vorhandenen geordneten Liste derjenige Flugstreifen entnommen, der für den momentanen Entscheidungsprozess von Interesse ist und auf dem Platz `next_Vehicle` abgelegt[48]. Die verbleibende Liste wird wieder auf dem Platz `Stripbay` abgelegt. Wurde das Luftfahrzeug ausgewählt, können die zugehörigen Informationen vom Flugstreifen extrahiert werden und die visuelle Aufmerksamkeit wird zurückgegeben, wobei hierbei wiederum eine Verzögerung von 3,9 Sekunden für den Blick auf den Flugstreifen addiert wird **(Annahme 8)**. Die visuelle Aufmerksamkeit ist demnach für 3,9 Sekunden für diesen Prozess gebunden (Pinska, 2008). Die extrahierte Information wird im Anschluss auf dem Platz `Visual_Buffer` abgelegt.

Durch das Ablegen aller aufgenommenen Informationen in den Plätzen `Visual_Buffer` bzw. `Auditory_Buffer`, werden diese dem nachgeschalteten Working-Memory-Modul zur Verfügung gestellt, welches in den Abschnitten 5.4.10 und 5.4.11 behandelt wird (vgl. auch Kapitel 3.4). Durch das Auftreten der Transition `look_at_flightstrip` wird außerdem die Planung initiiert, welche der folgende Abschnitt behandelt.

5.4.4 Phase der Planung

Die Phase der Handlungsplanung ist im MATriCS-Modell sehr stark anwendungsbezogen und reduziert umgesetzt, da der Hauptfokus des Modells auf den kognitiven Ressourcen und dem Arbeitsgedächtnis liegt.[49].

Das Planungsmodul der vorliegenden Version 1.9 (siehe Abbildung 30) beschreibt die Entnahme der Flugstreifen aus dem Drucker und anschließende Sortierung und Einordnung in die Stripbay. Die Planung wird hierbei in Anlehnung an Smieszek et al. (2011) global als Erstellung eines Staffelungs- bzw. Verkehrsplanes betrachtet. Das Modul beschreibt demnach

[48] Dies erfolgt über die Standard-ML-Funktion `List.filter (fn(v)=> #id(v)=ident) flightstriplist` am output-arc der Transition `choose_next_vehicle`. .

[49] Es ist allerdings jederzeit möglich, beispielsweise für die Untersuchung anderer Schwerpunkte, das Sub-Modul sehr viel detaillierter zu modellieren. Beispielsweise wäre die Planung weniger anwendungsbezogen und allgemeiner als kognitive Aufgabe der Koordination von Handlungszielen umsetzbar.

die Einordnung der neu erhaltenen, sowie der vom Lotsen selbst markierten Flugstreifen und deren Sortierung anhand ihrer geplanten Ankunftszeit in die Stripbay. Der Erhalt eines Flugstreifens triggert hier, wie bereits im vorherigen Kapitel beschrieben, die Planungsphase. Hierzu müssen zunächst die Informationen, die der Flugstreifen enthält gelesen und mental repräsentiert werden. Dabei wird ein Chunk im Arbeitsgedächtnis erstellt. Es wird damit die Enkodierung der Informationen beschrieben, die für den Planungsprozess nötig sind. Konkret bedeutet dies, dass der Lotse die geplante Ankunfts- oder Abflugszeit vom Flugstreifen abrufen muss, um zu wissen, an welcher Stelle der Stripbay der Flugstreifen einzusortieren ist **(Annahme 9)**.

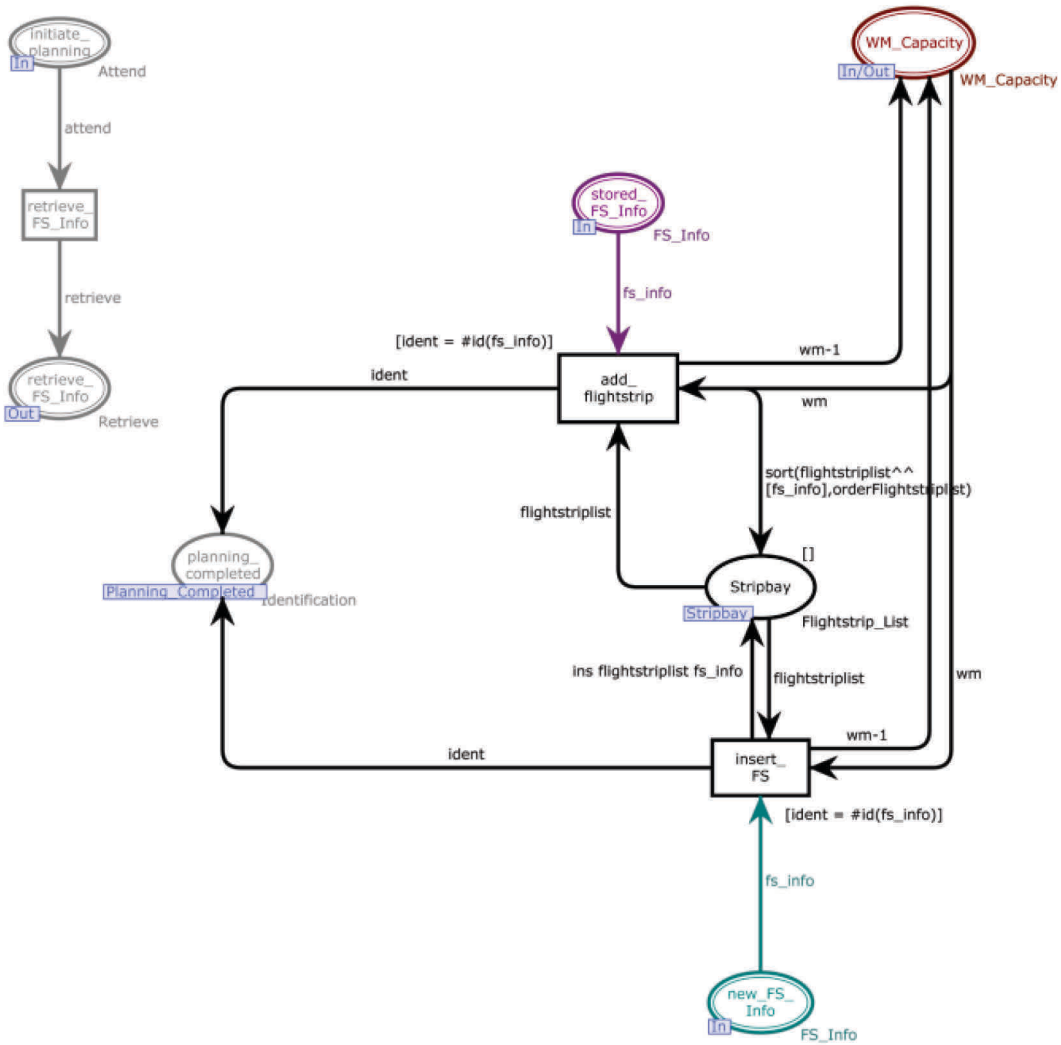

Abbildung 30: Das Planungsmodul im Petrinetz

5.4.5 Das plan-Modell im Petrinetz

Das Plan-Sub-Modul ist aufgrund dessen, dass der Fokus der Arbeit auf anderen Aspekten der Lotsenarbeit liegt, dementsprechend simpel gestaltet. Es besteht aus sieben Plätzen und drei Transitionen, wobei nochmals zwei unterschiedliche Prozesse abgebildet sind. Die Initiierung der Planung erfolgt wie beschrieben im `Perceive`-Modell. Hierdurch wird eine Marke des

Colorsets UNIT auf den Platz initiate_planning abgelegt, wodurch die Transition retrieve_FS_info aktiviert wird. Diese bewirkt, dass die Informationen des Flugstreifens aus dem Arbeitsgedächtnis abgerufen werden (vgl. auch Abschnitt 5.4.11). Da angenommen wird, dass ein Chunk eines Lotsen alle wichtigen Informationen zu einem Flugzeug enthält, also alle, die auf einem Flugstreifen (sprich im colorset Vehicle) enthalten sind, werden auch alle Informationen im Petrinetz repräsentiert, obwohl für die in diesem Sub-Modul abgebildete Aufgabe nur die geplante Ankunftszeit von Interesse wäre. Dies bringt ebenso den Vorteil einer einfacheren Umsetzung mit sich: Die Kommunikation mit dem Interaktionsmodell gestaltet sich hierdurch sehr viel einfacher, wenn ein Colorset verwendet wird (hier FS_Info), welches dem Colorset Vehicle gleicht, als wenn zwischen mehreren Colorsets mit verschiedenen Dateninhalten gewechselt werden müsste.

Durch den Abruf der Information aus dem Arbeitsgedächtnis wird eine Marke auf dem Platz stored_FS_info erstellt, wodurch die Transition add_flightstrip aktiviert wird, vorausgesetzt, die Marke auf dem Platz WM_Capacity gibt an, dass Arbeitsgedächtniskapazität zur Verfügung steht[50], um den Prozess kognitiv auszuführen[51]. Durch das Feuern der Transition wird der bereits auf dem Platz Stripbay enthaltenen Liste das Element auf dem Platz stored_FS_info hinzugefügt[52]. Das eigentliche Sortieren anhand der geplanten Ankunftszeit (Variable s_time) wird anschließend auf die um das neue Element ergänzte Liste angewendet. Dies wird durch mehrere Standard-ML Funktionen (fun insert; fun sort; fun orderFlightstriplist) übernommen, welche am Output-Arc der Transition add_filghtstrip definiert sind. Diese erzeugen dann die geordnete Liste aller Flugstreifen auf dem Platz Stripbay. Zusätzlich wird auf dem Platz planning_completed eine Marke erzeugt, welche Voraussetzung dafür ist, dass eine Top-Down getriggerte Aufmerksamkeitsverschiebung zur Stripbay erfolgen kann[53].

5.4.6 Phase der Entscheidungsfindung

Entscheiden beschreibt den Prozess des Auswählens aus mehreren Alternativen z.B. die Entscheidung für eine bestimmte Vorgehensweise (Roske-Hofstrand & Murphy, 1998). Dies

[50] Die Marke auf dem Platz WM_Capacity kann Zahlenwerte zwischen 0 und 10 (Anzahl der Chunks) annehmen, wobei bei einem Wert von 10 keine weiteren Chunks im Arbeitsgedächtnis speicherbar sind.
[51] In der Version 1.8 sind für das Hinzufügen des Flugstreifens motorische Handlungen auszuführen, die durch den Wegfall des Planungsschrittes im Validierungsexperiment in Version 1.9 entfernt wurden.
[52] Initial ist diese Liste leer.
[53] Als weitere Vereinfachung wurde die Stripbay im „Kopf des Lotsen" modelliert – wohl wissend, dass sie ein externes Hilfsmittel darstellt. Da sie aber nicht dem Informationsaustausch zwischen Lotsen und Flughafen dient, wurde sie nicht im Interaktionsmodell modelliert, da die Erstellung der nötigen Verbindungen zwischen Lotsen- und Interaktionsmodell sehr komplex wären. Die Auswirkungen auf die Performanz (Arbeitsgedächtnisbelegung, manuelle Handlungen) sind allerdings bei beiden Varianten der Umsetzung gleich.

kann in der Flugverkehrskontrolle beispielsweise die Entscheidung sein, welchem Luftfahrzeug zuerst eine Freigabe erteilt werden sollte. Dieser Prozess ist oft mit Unsicherheiten verbunden und kann von individuellen Einstellungen, Erwartungen und kognitiven Tendenzen (Biases) beeinflusst werden. Roske-Hofstrand und Murphy (1998) beschreiben, dass es in der Flugverkehrskontrolle wahrscheinlich ist, dass Entscheidungen anhand von Regeln getroffen werden (Entscheidungsheuristiken), die in der Vergangenheit gut funktionierten (vgl. Kahnemann, Slovic, & Tversky, 1982; Wickens, 1992). Die Effizienz und Effektivität der Entscheidungen wird dabei unter Anderem von der erforderlichen Anstrengung, der zur Verfügung stehenden Zeit im Vergleich zur benötigten Zeit, den Erwartungen sowie Begrenzungen der Aufmerksamkeit und des Arbeitsgedächtnisses eingeschränkt. Simon (1955) beschreibt, dass Menschen oft möglichst schnell, mit begrenztem Wissen und begrenzter Informationsverarbeitung Entscheidungen treffen müssen. In klassischen Entscheidungs-situationen wenden Menschen daher Entscheidungsheuristiken an, um ihren begrenzten Informationsverarbeitungsressourcen Rechnung zu tragen (Gigerenzer et al., 1999; Simon, 1955, 1956). Solche Heuristiken ignorieren häufig die Regeln der Wahrscheinlichkeit und wurden daher lange Zeit als Verzerrungen (Biases) oder Abweichungen von rationalen Entscheidungen angesehen. Moray (1986) beschreibt 25 solcher Biases, welche allerdings tatsächlich adaptive Strategien zum Umgang mit Kapazitätsbegrenzungen darstellen.

Hollnagel (2012) beschreibt das Treffen von Entscheidungen als eine Ausprägung des ETTO-Prinzips (vgl. Kapitel 1). Bereits Lindblom (1959) beschreibt das kognitive Verhalten von Menschen als eine Reihe von Trade-Offs. Um mit diesen Trade-Offs umgehen zu können, legen Menschen ein Verhalten an den Tag, das er mit „Muddeling-Through“ („sich durchwursteln“) umschreibt (vgl. auch Dörner, 2000, S. 84; Hollnagel, 1992). Entscheidungen können daher in vier Schritte unterteilt werden: (1) Definiere das prinzipielle Ziel. (2) Umreiße ein paar offensichtliche Alternativen. (3) Wähle eine Alternative aus, die einen vernünftigen Kompromiss zwischen Kosten und Nutzen darstellt. (4) Wiederhole das Vorgehen, wenn das Resultat unbefriedigend ist (Hollnagel, 2012). Das gezeigte Verhalten wird hierbei allerdings nicht mit begrenzter Informationsverarbeitungskapazität erklärt, sondern durch eine Anpassung an die Umgebungsbedingungen. Einigkeit besteht allerdings darüber, dass Menschen nicht nach optimalen Entscheidungen streben, sondern dass Entscheidungen eher so getroffen werden, dass sie den Ansprüchen genügen (Simon, 1955). Ein solches Vorgehen bei Entscheidungen wird auch mit *Naturalistic Decision Making* bezeichnet (Klein, 2008; Todd & Gigerenzer, 2001; Zsambok & Klein, 1997). Möhlenbrink (2011) untersucht in diesem Zusammenhang, inwiefern Fluglotsen Entscheidungsheuristiken anwenden, um den Flugverkehr effizient zu bearbeiten. Er implementiert verschiedene solcher Heuristiken in ein holistisches Flughafen-Lotsenmodell (vgl. Werther, 2006a), um diejenigen Handlungsstrategien abzubilden, die Probanden in einer

experimentellen Untersuchung in einer Mikrowelt prinzipiell verfolgen können. Hierbei implementierte er sowohl die einfachste „first come, first served"-Heuristik, als auch einstufige Hierarchien (z.B. Land > TakeOff > TaxiIn > Push > TaxiOut > TaxiInA) und zweistufige Hierarchien (z.B. (TaxiIn > Land > TakeOff) > (TaxiOut > TaxiInA > Push)) (vgl. Möhlenbrink, 2011).

Die starke Reglementierung der Flugverkehrskontrolle sieht zusätzlich für viele Eventualitäten und komplexitätserhöhende Faktoren sogenannte Standard Operating Procedures (SOPs) vor. Aus Koros et al. (2006) geht hervor, dass Lotsen hauptsächlich nach diesen SOPs handeln und ihre Entscheidungen danach treffen. Eine situationsbedingte Entscheidung muss der Lotse beispielsweise treffen, wenn zwei Luftfahrzeuge gleichzeitig eine Flughafen-Ressource (z.B. die Start-/Landebahn) verwenden wollen. Hierfür besagt die SOP beispielsweise, dass ankommende Flüge immer Vorrang vor abfliegenden haben sollten (vgl. Kapitel 4.4.4). Die Schwierige Aufgabe des Lotsen ist es dabei, zu evaluieren, ob und wie viel Zeit besteht, um beispielsweise zwischen zwei ankommenden Flügen einem abfliegenden die Starterlaubnis zu erteilen. Eine solche Evaluation ist bisher noch nicht im MATriCS-Modell umgesetzt worden.

Da der Fokus des hier vorliegenden Modells auf der Beschreibung der kognitiven Ressourcen Aufmerksamkeit und Arbeitsgedächtniskapazität liegt und weniger auf der Implementierung adäquater Entscheidungsalgorithmen oder Heuristiken, wurde die einfache „first come, first served"-Heuristik nach Möhlenbrink (2011) implementiert **(Annahme 10)**. Es wird dabei davon ausgegangen, dass alle Informationen, die in den vorangegangenen Phasen aufgenommen, gespeichert und bearbeitet worden sind, im Entscheidungsmodell zusammenlaufen. Aus der Annahme, dass ein Lotse allein für einen Flughafen zuständig ist, ergibt sich außerdem eine weitere Unterteilung des Entscheidungsprozesses in drei Unterprozesse, welche zunächst als parallel ablaufend angesehen werden **(Annahme 11)**: Das Lotsenmodell kann auf Basis seiner Informationen entscheiden ob (1) eine Start- oder Landefreigabe gegeben werden muss/kann (`Location_Free_Landing _and_ Takeoff`); (2) eine Rollfreigabe gegeben werden muss/kann (`Location_Free_ Taxi`) oder (3) eine Freigabe zum Abrollen von der Start-/Landebahn hin zur Parkposition gegeben werden kann/muss (`Location_Occupied`) (siehe Abbildung 31 auf Seite 113). Dieser Entscheidungsprozess basiert dabei ausschließlich auf der momentanen Situation und dem momentanen Status der betreffenden Location. Für Entscheidungen, die die Start-/ Landebahn betreffen ist dieser Prozess ausreichend. Der Zustand der Piste (frei; belegt) wird ständig im mentalen Modell des Lotsenmodells repräsentiert und wenn nötig durch das Auftreten von Handlungen oder externen Ereignissen aktualisiert.

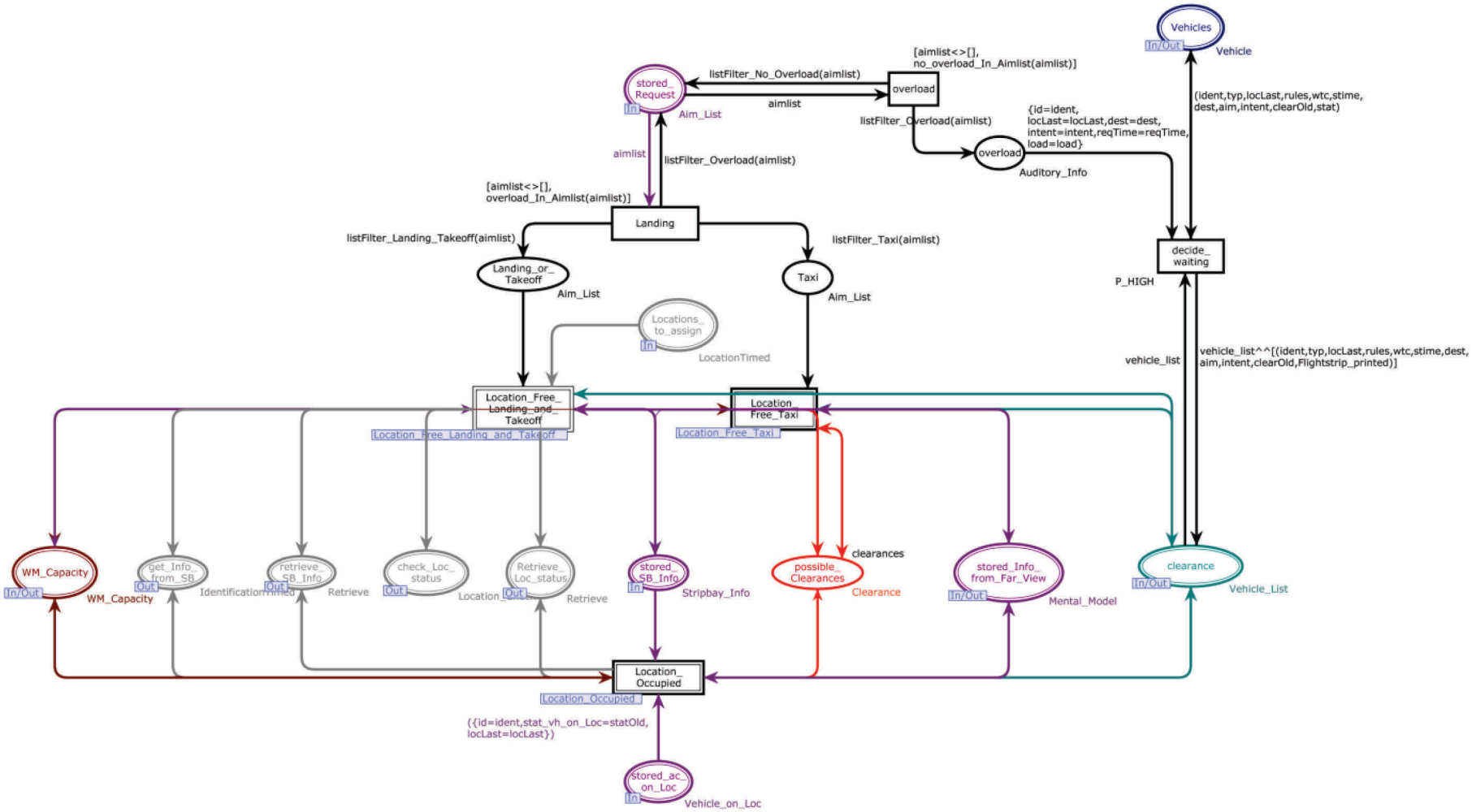

Abbildung 31: Das Entscheidungs-Modul im Petrinetz

Entscheidungen, die das Rollen betreffen, müssen allerdings den Zustand aller Rollwegsabschnitte der angefragten Rollroute beinhalten. Die Entscheidungen des Lotsen müssen außerdem nicht nur die aktuellen Zustände dieser Abschnitte, sondern auch deren mögliche zukünftige Zustände berücksichtigen. Bereits aus den Aufgabenanalysen geht hervor, dass der Lotse die Situation in die nahe Zukunft projizieren, sie mental simulieren muss (Klein & Crandall, 1995; Klein, Moon, & Hoffmann, 2006, vgl. Kapitel 4.3.7 und 4.5). Die theoretischen Grundlagen für die Implementierung eines solchen Entscheidungsmodells beschreiben Smieszek und Joeres (2013). Sie identifizieren aber auch, dass zwar die mentalen Prozesse, die bei der Schätzung von Zeitintervallen eine Rolle spielen, recht gut untersucht sind (vgl. Rußwinkel, Urbas, & Thüring, 2011), jedoch ist die Übertragung dieser Prozesse auf die Projektion von Routen noch nicht ausreichend erforscht. Ferner ist es ein ebenso wenig triviales Unterfangen zu bestimmen, wie hoch der Bedarf an kognitiven Ressourcen für eine solche mentale Simulation ist. Aus diesen Gründen wurden die Abläufe, die für Entscheidungen, die die Start-/Landebahn betreffen, auch für Entscheidungen über die Rollwege übernommen **(Annahme 12)**, wohl wissend, dass hierbei sehr viel komplexere Prozesse involviert sind (vgl. Smieszek & Joeres, 2013). Die kognitive Beanspruchung ist dabei lediglich von der Anzahl der zutreffenden Entscheidungen und damit der hierfür aufzunehmenden und zu speichernden Informationen abhängig, nicht aber von der Komplexität der jeweiligen Entscheidung[54] und dem mit dem mentalen Simulationsprozess verbundenen Ressourcenbedarf **(Annahme 13)**.

5.4.7 Das Entscheidungsmodell im Petrinetz

Das Entscheidungsmodell besteht aus 16 Plätzen und sechs Transitionen, wobei auf den darunterliegenden Ebenen nochmals jeweils zwölf Plätze und zwei Transitionen hinzukommen. Grundlegend wird davon ausgegangen, dass jeder der Entscheidungsabläufe durch die Anfrage eines Piloten getriggert wird. Dabei wird immer diejenige Freigabe zuerst abgearbeitet, welche zuerst den Lotsen erreicht hat („first come, first served"-Heuristik anhand der Variable `reqTime`). Es können allerdings gleichzeitig Entscheidungsprozesse, die Start und Landung und das Rollen betreffen, getriggert werden. Für das Treffen einer Entscheidung werden weitere Informationen benötigt: der korrespondierende Flugstreifen aus der Stripbay, sowie die Information über den Zustand der jeweiligen zu verwendenden Location durch einen Blick nach draußen (vgl. auch Kapitel 5.4.2). Durch die Begrenzung der kognitiven Ressourcen (vor allem visuelle Aufmerksamkeit) wird ein serielles Abarbeiten zweier gleichzeitig zu treffender

[54] Prinzipiell ist anzunehmen, dass eine Entscheidung über eine Freigabe für ein Flugzeug sehr viel einfacher, mit geringerem kognitivem Aufwand, zu treffen ist, wenn es sich allein auf dem Flughafen befindet, als wenn mehrere andere Flugzeuge die gleichen Rollwegsabschnitte zu bestimmten Zeitpunkten ebenfalls nutzen wollen. Hierbei ist zu bestimmen, wann welche Rollwegsabschnitte belegt sein werden, wann es potentiell zu Kollisionen kommen kann und ob eventuell nur eine teilweise Freigabe gegeben werden sollte (vgl. Kapitel 4.5).

Entscheidungen erreicht. So kann immer nur eine Information gleichzeitig visuell enkodiert werden, während die visuelle Aufmerksamkeit für einen Prozess gebunden ist. Ferner können bestimmte Entscheidungsprozesse nicht abgeschlossen werden, solange die benötigte Location (z.B. die Start-/Landebahn) belegt ist (es wird z.B. keine Freigabe erteilt, solange die Start-/Landebahn belegt ist). Wenn dies der Fall ist, muss zunächst der Prozess `Location_Occupied` ablaufen, um die Location wieder frei zu machen bzw. die Nachfolgefreigabe (das Rollen zur Parkposition) für das okkupierende Luftfahrzeug zu erteilen. Über das Perceive-Modul und das Arbeitsgedächtnis-Modul werden dabei die eingehenden Anfragen aus dem Flughafenmodell in das Entscheidungsmodul `decide` weitergeleitet. Es werden alle zu einem Zeitpunkt anstehenden Anfragen auf dem Platz `stored_Request` abgelegt. Diese Anfragen können dabei als Unterziele des Lotsen verstanden werden. Durch die Funktionen `listFilter_Landing_Takeoff(aimlist)` und `listFilter_Taxi (aimlist)` an den Output-Arcs der Transition `split_requests` werden diese anschließend in Anfragen zum Starten und Landen, sowie für das Rollen aufgeteilt. Hierbei entsteht jeweils eine anhand der Variable `reqTime` geordnete Liste aller Anfragen auf den nachgeschalteten Plätzen `Landing_or_Takeoff` und `Taxi`. Diese stehen dann jeweils für die gezielten Entscheidungsprozesse in den Substitutions-Transitionen `Location_Free_Landing_ and_Takeoff` (siehe Abbildung 32 auf Seite 116) und `Location_Free_Taxi` zur Verfügung. Beide Sub-Module sind dabei ähnlich aufgebaut, da sich, wie oben beschrieben, das Entscheidungsmodell für Rollfreigaben am Modell für Start- und Landefreigeben orientiert. Bei Vorliegen einer Anfrage sind zunächst verschiedene Bedingungen zu überprüfen bzw. Informationen zu sammeln, um zu entscheiden, ob eine Freigabe gegeben werden kann. Hierzu feuert zunächst die Transition `check_constraints`, wenn die notwendigen Eingangsbedingungen erfüllt sind. Für die gegebene Intention des Luftfahrzeuges muss daher zunächst eine entsprechende Freigabe vorgesehen sein. Das bedeutet, dass in der Marke, welche die Anfrage beschreibt, für die aktuelle Location, die Ziellocation und die gegeben Intention des anfragenden Luftfahrzeuges eine entsprechende Freigabe auf dem Platz `possible_ Clearances` vorhanden sein muss. Dieser Platz kann als Teil des Langzeitgedächtnisses[55] des Lotsen angesehen werden, in dem Informationen darüber abgelegt sind, für welche Intentionen von welchen Ausgangspunkten prinzipiell welche Freigaben erteilt werden können.

Für die Erteilung von Start- und Landefreigaben muss außerdem gewährleistet sein, dass eine Marke für die entsprechende Start-/Landebahn auf dem Platz `Locations_to_assign` vorhanden ist. Dieser Platz kann

[55] Ein Langzeitgedächtnis wurde nicht explizit modelliert, sodass hier diese beiden beschriebenen Plätze stellvertretend die Informationen enthalten, die prinzipiell in einem Langzeitgedächtnis gespeichert wären.

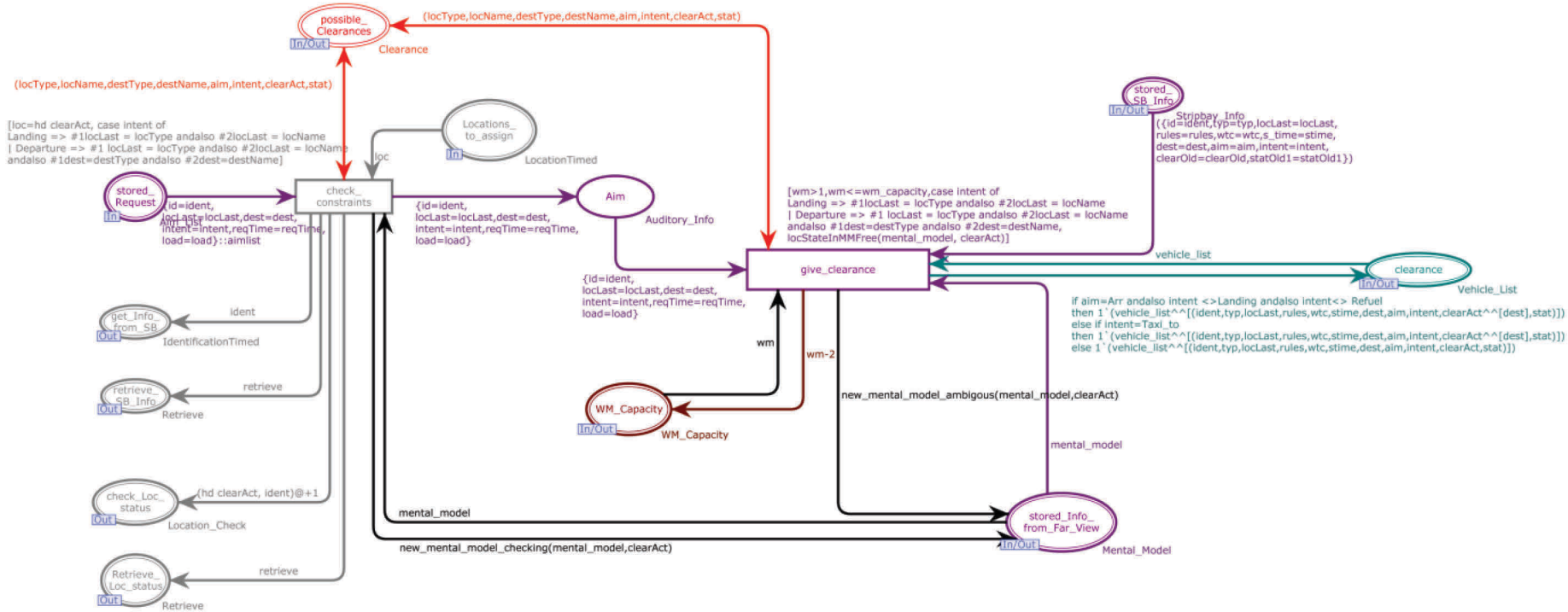

Abbildung 32: Entscheidungs-Sub-Modul für Starts- und Landungen im Petrinetz

ebenfalls als Teil des Langzeitgedächtnisses angesehen werden, welcher Informationen darüber enthält, welche Start-/Landebahnen prinzipiell existieren. Durch das Feuern der Transition `check_constraints` werden eine Reihe Marken auf den nachgeschalteten Plätzen generiert, welche einerseits die Top-Down Trigger für die aktive Informationsaufnahme darstellen, andererseits die Voraussetzungen für den Abruf dieser Informationen aus dem Arbeitsgedächtnis darstellen (vgl. Kapitel 5.4.3 und 5.4.11). Auf dem Platz `get_Info_from_SB` wird eine Marke abgelegt, welche die Aufmerksamkeitsverschiebung zur Stripbay gewährleistet, um die Informationen des Flugstreifens abzurufen. Die Aufmerksamkeitsverschiebung zur Außensicht wird durch Ablage einer Marke auf dem Platz `check_loc_status` initiiert, welche im Weiteren zum Abruf der Informationen über den Zustand der Location dient (vgl. Kapitel 5.3). Die Generierung jeweils einer Marke auf den Plätzen `retrieve_SB_Info` und `retrieve_Loc_Status` dient der Vorbereitung des Abrufes der aufgenommenen Informationen aus dem Arbeitsgedächtnis. Sind diese Informationsaufnahme- und Abrufprozesse abgeschlossen, so wird auf den Plätzen `stored_SB_info` und `stored_Info_from_Far_View` jeweils eine Marke abgelegt, die die betreffenden Informationen des Flugstreifens bzw. über den Zustand der Location (free oder occupied) enthält. Ist der Zustand der Location „free“, sowie der korrespondierende Flugstreifen abgerufen, so wird die Entscheidung gefällt, dass eine Freigabe erteilt werden kann. Über die Funktion `locStateInMMFree(mental_model, clearAct)` an der Transition `give_clearance` wird hierbei die Voraussetzung überprüft, dass die benötigte Location auf dem Platz `stored_Info_from_Far_View` den Status „free“ enthält. Ist dies der Fall, so kann die Transition feuern. Dabei werden alle zuvor aufgenommenen Informationen integriert, um diese Entscheidung zu generieren. Durch das Feuern der Transition `give_clearance` wird eine Marke, welche die Freigabe enthält, auf dem Platz `clearance` erstellt.

Ist die betreffende Location auf dem Platz `stored_Info_from_Far_View` mit dem Status „occupied“ versehen, so kann die Transition nicht feuern, da zunächst der Prozess `location_occupied` ablaufen muss. Dieser wird im Anschluss an die Informationsaufnahme aus der Außensicht getriggert. Das Modell stellt hierbei fest, dass die Location (z.B. die Piste) noch von einem anderen Luftfahrzeug belegt ist (Platz `vehicle_on_loc` im Interaktionsmodell; vgl. Kapitel 5.3), welches zunächst eine Freigabe erhalten muss, damit die Piste wieder frei ist. Hierbei wird der erhaltene Trigger im Arbeitsgedächtnis gespeichert und steht dem Entscheidungsprozess dann auf dem Platz `stored_ac_on_Loc` zur Verfügung. Für die Ausführung dieser Entscheidung müssen nur noch die Informationen des zugehörigen Flugstreifens abgerufen werden, da die Information über die Belegung der Location bereits vorhanden ist. Wurde die Flugstreifeninformation abgerufen, so kann die Transition

give_clearance feuern, womit die Freigabe zum Abrollen von der Piste und zum Rollen zur Parkposition erteilt wird. Es wird ebenso eine Marke auf dem Platz clearance erstellt. Weiterhin wird durch das Treffen von Entscheidungen jeweils Kapazität im Arbeitsgedächtnis freigegeben (Platz WM_Capacity). Die freigegebene Kapazität richtet sich dabei danach, welche Aktionen im Anschluss noch ausgeführt werden müssen und welche Informationen daher noch weiter im Arbeitsgedächtnis aufrechterhalten werden müssen (**Annahme 14**; Zeigarnik-Effekt; Zeigarnik, 1927). Beispielsweise muss bei einer Rollanfrage zur Piste nach der Erteilung der Freigabe der Flugstreifen markiert und wieder in die Stripbay einsortiert werden, wobei so lange Informationen aufrecht erhalten werden müssen, bis dieser Prozess abgeschlossen ist[56]. Die auf die Entscheidung folgenden Prozesse sind im nachfolgenden Abschnitt erläutert.

5.4.8 Phase der Handlungsausführung

Hat der Lotse eine Entscheidung getroffen, so muss er sie an den Piloten übermitteln. Die geschieht über die Funkkommunikation. Zusätzlich müssen weitere Handlungen ausgeführt werden, wie beispielsweise das Markieren der Flugstreifen, das Entnehmen des Flugstreifens oder die Weitergabe des Flugstreifens an die nächste Kontrollposition (Handover). Neben der Aufnahme und der Verarbeitung von Informationen muss der Lotse motorische Bewegungen ausführen. Zusätzlich zu den kognitiven Prozessen und den extern ablaufenden Verkehrsprozessen werden auch diese motorischen Prozesse und die damit verbundenen Zeitverzögerungen im Modell abgebildet. Die Zeit, die ein Lotse für die von ihm durchgeführten Bewegungen benötigt, lässt sich dabei auf zwei Prozesse aufteilen: Zum einen müssen motorische Prozesse vorbereitet und initiiert, zum anderen physisch ausgeführt werden **(Annahme 15)**. Die Dauer der *physischen Bewegungen* lassen sich anhand von Fitts' Law (Fitts, 1954) approximieren **(Annahme 16)**. Da es sich bei den hier betrachteten Bewegungen von Lotsen um Bewegungen im dreidimensionalen Raum handelt, etwa das Greifen nach oder das Ablegen von Flugstreifen, wird zur Abschätzung der Dauer dieser Greifbewegungen die allgemeine Form von Fitts' Law nach Welford (Welford, 1960, 1968; vgl. McKenzie, 1992) angewendet:

$$T = a + b \cdot \log_2\left(\frac{A}{W} + 0{,}5\right)$$

Dabei beschreibt A die relative Entfernung, die bei einer Bewegung zurückgelegt wird. W beschreibt die relative Zielbreite (vgl. Abbildung 33).

[56] Hierbei sei nochmals darauf hingewiesen, dass in der aktuellen Version kein zeitlicher Decay von Informationen implementiert wurde, die Informationen also solange bestehen bleiben, bis sie „aktiv" vergessen werden.

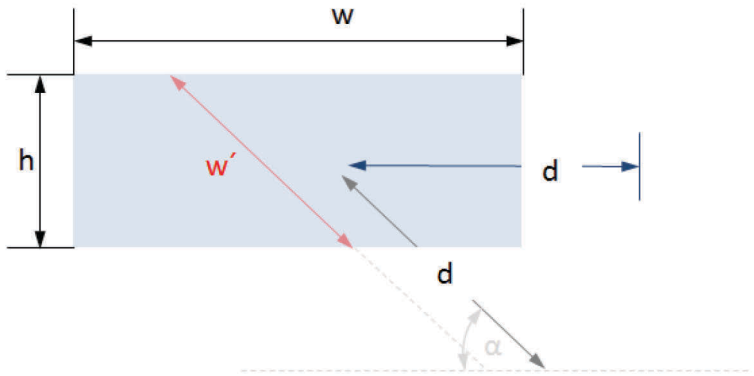

Abbildung 33: Relative Zielbreite (eigene Darstellung nach Preim und Dachselt, 2010)

Diese ist abhängig von der Richtung, aus der ein Ziel angesteuert wird. In diesem Fall ist die relative Zielbreite minimal, wenn das Zielfeld mit einem Winkel von $\alpha = 90°$ angesteuert wird und maximal, wenn $\alpha = arctan(h/w)$ *ist.* Für dreidimensionale Armbewegungen wurden die Koeffizienten a und b experimentell von Kerr und Langolf (1977) ermittelt, die die folgende Gleichung aufstellen:

$$T = 78 \cdot \log_2\left(\frac{A}{W} + 0{,}5\right) - 10ms$$

Um alle möglichen Bewegungen zu beschreiben, müssen also die *Entfernung* und die *relative Zielbreite* zwischen jedem motorischen Startpunkt A und Zielpunkt B bestimmt werden. Zu diesem Zweck werden *motorische Locations* definiert, die der Lotse im Laufe seiner Arbeit erreichen muss. Diese Locations basieren dabei in der vorliegenden Modellversion nicht auf dem realen Arbeitsplatz eines Towerlotsen, sondern auf dem experimentellen Setting (vgl. Kapitel 6.1.2), in dem das Gesamtmodell validiert wurde. Diese können aber jederzeit an das Layout realer Towerarbeitsplätze angepasst werden. Hierzu kann die im Anhang befindliche Datei „Zeitberechnung_manual_Welford_v1.8.xlsx“ verwendet werden.

Aus dem Arbeitsablauf des Towerlotsen, bzw. der Versuchspersonen im Experiment, ergeben sich fünf Locations:

- Ein Bereich, in dem die unterschiedlichen Stifte für verschiedene Aufgabenteile liegen (*pens/default*).
- Die Stripbay (*SB*), die zur Aufnahme und zur Markierung der Flugstreifen angesteuert wird.
- Der Ablagekorb, in den die Flugstreifen abgearbeiteter Flugzeuge gelegt werden (*HO*).
- Die Sprechtaste des Funkgerätes (*Radio*).
- Der Fragebogen zur Erhebung der momentanen mentalen Beanspruchung (*rsme*)[57].

[57] Die hier definierten motorischen Loctaions basieren auf dem Arbeitsplatz der Versuchspersonen im Validierungsversuch. Für einen realen Lotsenarbeitsplatz wurden in MATriCS-Version 1.8 folgende Locations definiert: Stripbay/Default, Flugstreifen-Drucker (FS-Printer), Leseposition für Flugstreifen (Read), Funkgerät (Radio) und Handover.

Jede dieser Locations wird als Quader definiert, in den die Hand hineinbewegt werden muss. Die Hand wird dabei vereinfacht als geometrischer Punkt betrachtet. Diese Quader liegen in einem Koordinatensystem, dessen Ursprung in der vorderen linken Kante des Arbeitstisches liegt. Die y-Achse zeigt dabei in Blickrichtung des Lotsen, die x-Achse (vom Lotsen aus) nach rechts und die z-Achse nach oben (vgl. Abbildung 34).

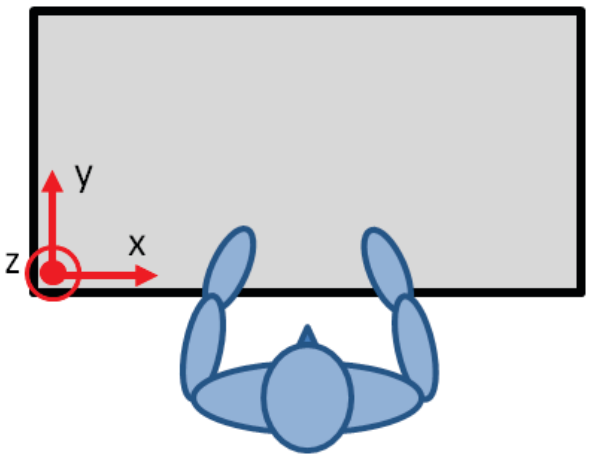

Abbildung 34: Koordinatensystem am Lotsenarbeitsplatz

Jede Location wird durch zwei Vektoren beschrieben. Der erste Vektor ($\overrightarrow{POS_A}$) beschreibt dabei die Position des Quaders, der zweite Vektor ($\overrightarrow{SIZE_A}$) beschreibt die Ausdehnung in den drei Koordinatenrichtungen. Bei fünf Locations ergibt sich eine Anzahl von $n_{mov} = 2 \cdot (4 + 3 + 2 + 1) = 20$ möglichen Bewegungen. Dabei ist zu beachten, dass Bewegungen zwischen den gleichen Locations aber mit umgekehrter Richtung getrennt betrachtet werden müssen ($A \rightarrow B \neq B \rightarrow A$). Es ergibt sich also für jede mögliche Bewegung von einem beliebigen Punkt A zu einem beliebigen Punkt *B:*

- Für den relativen Abstand: $A = |\overrightarrow{POS_B} - \overrightarrow{POS_A}|$
- Für die relative Zielbreite: $W = 2 \cdot \left| \overrightarrow{POS_B} - \overrightarrow{S_{AB}} \right|$

Wobei $\overrightarrow{S_{AB}}$ der Schnittpunkt der Geraden durch $\overrightarrow{POS_B}$ und $\overrightarrow{POS_A}$ mit dem Quader um die Location *B* ist. Für die Berechnung der Bewegungsdauer wurde die im Anhang befindliche Datei „Zeitberechnung_manual_Welford_v1.9.xlsx“ verwendet.

Für die Dauer der *Bewegungsvorbereitung* werden an dieser Stelle die Berechnungsmechanismen der kognitiven Architektur ACT-R zugrunde gelegt (Bothell, 2007). Hier wird angenommen, dass die Vorbereitungsdauer einer Bewegung von ihrer Komplexität abhängig ist. Die Komplexität ist wiederum abhängig von der Anzahl der *Features*, anhand derer die Bewegung definiert wird. Jedes Feature ist eine elementare Information über die entsprechende Bewegung, wie etwa die Bewegungsart, die verwendete Hand, evtl. verwendete Finger, sowie die Bewegungsrichtung und –weite **(Annahme 17)**.

Dabei werden fünf verschiedene Bewegungsarten unterschieden, die in Bothell (2007, S.309 f) beschrieben werden. Für jedes Feature, das zu einer Bewegung gehört, wird eine Vorbereitungsdauer von *50 ms* veranschlagt. Tabelle 2 zeigt die Features, die den fünf definierten Locations zugeordnet werden und die daraus resultierenden Vorbereitungszeiten. Zusätzlich zur jeweiligen Vorbereitungsdauer wird eine *Initiation*-Dauer von 50*ms* berücksichtigt.

Tabelle 2: Bewegungstypen, die damit verbundene Anzahl der Features und die sich daraus ergebende Vorbereitungszeit für die definierten Ziel-Locations

Bewegungstyp / Ziel-Location	**Notwendige Features**	**Vorbereitungs-dauer**
Griff nach einem neuen Stift (→ *pens*)	• Motion Style: Point-hand • Hand: right[58] • Direction • Distance • Motion Style: Peck [als Greifersatz] • Hand: right[45] • Finger: index • Finger: thumb ➔ 6 Features	250ms + 50ms
Bewegung zur Stripbay (→ *SB*)	• Motion Style: Point-hand • Hand: right[46] • Direction • Distance ➔ 3 Features	150ms + 50ms
Übergeben der FS (*HO*)	• Motion Style: Ply • Hand: right[46] • Direction • Distance • Motion Style: Peck [als Ersatz für Loslassen] • Hand: right[46] • Finger: index ➔ 5 Features	250ms + 50ms
Aktivieren des Funkgerätes (*radio*)	• Motion Style: Point hand • Hand: righ[46] • Direction • Distance • Motion Style: Peck • Hand: right[46] • Finger: index ➔ 5 Features	250ms + 50ms
Bewegung zum Fragebogen (*rsme*)	• Motion Style: Ply • Hand: right[46] • Direction • Distance ➔ 3 Features	150ms + 50ms

5.4.9 Das Handlungs-Modell im Petrinetz

Das `Act`-Modul besteht aus 18 Plätzen und sieben Transitionen, sowie nochmals fünf Plätzen und zwei Transitionen um `Manual`-Modul. Motorische Prozesse treten im Modell immer dann auf, wenn der Lotse Flugstreifen modifiziert oder weiter gibt, wenn er Fragebögen ausfüllt, oder den Sprechfunk aktiviert. Diese Operationen sind im Modell-Modul `act` angesiedelt (siehe Abbildung 35 auf Seite 122). Zeitlich können motorische Handlungen allerdings auch parallel zu den ablaufenden kognitiven Prozessen ausgeführt werden

[58] Die ausführende Hand und damit das entsprechende Feature ist immer dasselbe. Aus diesem Grund darf das Feature nicht mitgezählt werden.

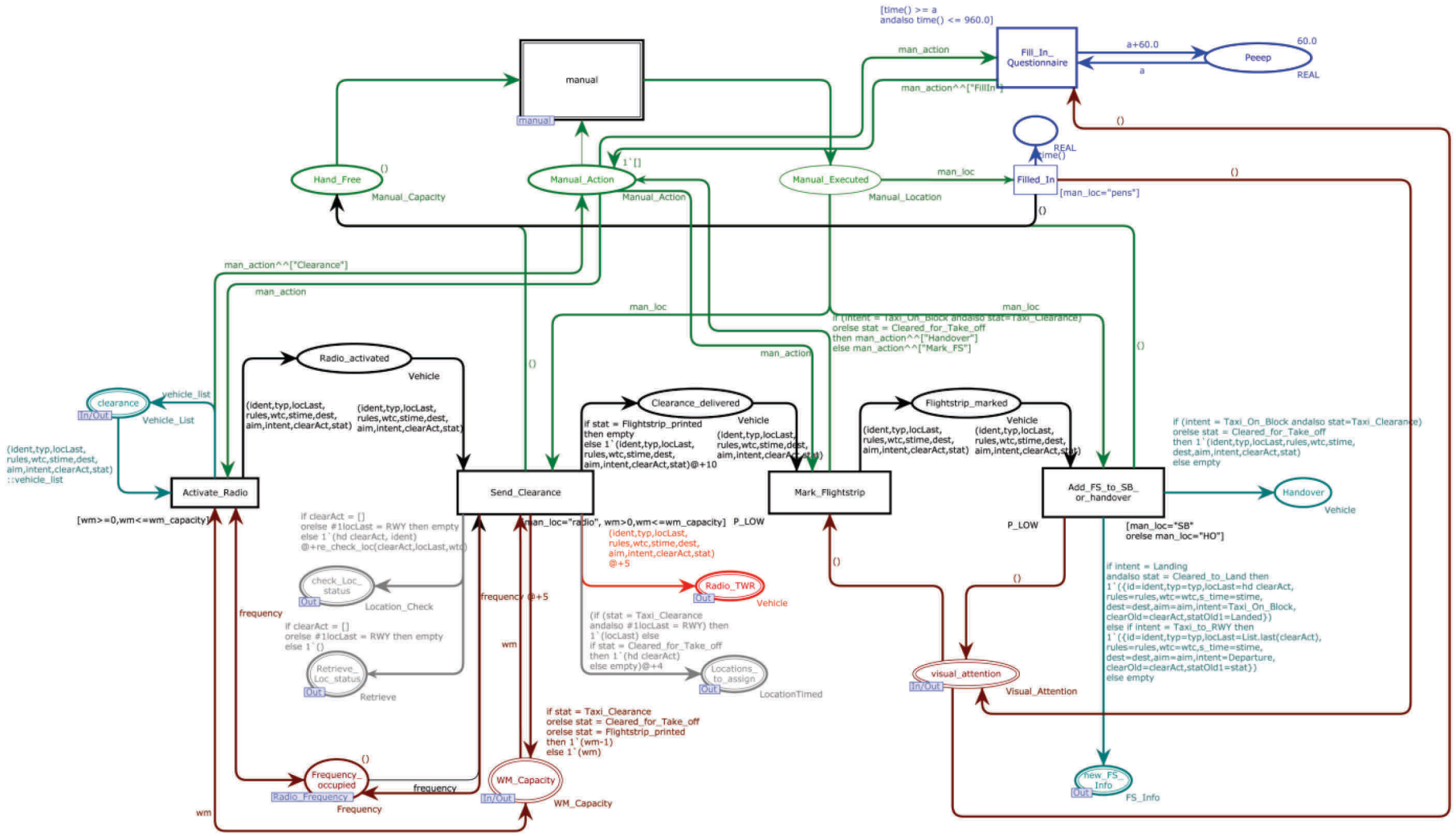

Abbildung 35: Das Act-Modul im Petrinetz

(Wickens, 1984; vgl. Kapitel 2.1; **Annahme 18**) und somit kann das act-Modul auch Trigger enthalten, die aus allen anderen Modulen stammen. Dies ist immer dann der Fall, wenn für die Ausführung kognitiver Aufgaben zunächst motorische Handlungen erforderlich sind (z.B. das Entnehmen von Flugstreifen aus dem Drucker[59]). Motorische Aktionen sind vorwiegend im Anschluss an die Entscheidung notwendig. Durch das Ablegen einer Marke auf dem Platz clearance werden dabei nacheinander diejenigen Prozesse aktiviert, welche motorische Handlungen erfordern. Diese sind das Aktivieren des Funkgerätes zur Übermittlung der Freigabe (Transition activate_Radio), die Markierung des Flugstreifens (Transition mark_Flightstrip) und das Ablegen des Flugstreifens in der Stripbay bzw. die Übergabe zur nächsten Kontrollposition (Transition Add_FS_to_SB_or_handover). Zusätzlich ist für die präzise Abbildung der Bedingungen des Validierungsexperimentes ein weiterer Trigger implementiert, welcher das Ausfüllen des Fragebogens modelliert. Hierbei wird durch den Platz peep (simuliert den Piepton, der die Versuchspersonen aufforderte, den Fragebogen auszufüllen) alle 60 Modellzeitschritte (also alle 60 Sekunden) die Transition Fill_in_Questionaire aktiviert, welche eine Anfrage für die entsprechende Bewegung an das manuelle Modul stellt.

Beim Schalten der genannten Transitionen werden auf dem Platz Manual_Action jeweils die zugehörigen Anfragen an das motorische Modul (siehe Abbildung 36) abgelegt, bevor sie (wie weiter unten erläutert) von diesem verarbeitet werden. Wird zurzeit keine motorische Operation ausgeführt, liegt auf dem Platz Hand_Free eine Marke des Colorset UNIT, welche anzeigt, dass das motorische Modul gerade frei ist. Der Platz Manual_Executed dient als Output-Kanal des motorischen Moduls an das übergeordnete act-Modul und zeigt an, wenn eine motorische Bewegung ausgeführt ist. Das Vollenden einer motorischen Aktion ist meist Voraussetzung für weitere Simulationsschritte.

Wird von einer der oben genannten Transitionen eine Anfrage gestellt, gelangt diese über den Platz Manual_Action an das motorische Modul. Derartige Anfragen werden in einer Liste auf dem gleichnamigen Platz abgelegt und umfassen eine komplette Aktion, wie z.B. „Mark_FS" (markieren eines Flugstreifens). Diese wird im motorischen Modul in einzelne Teilbewegungen aufgespalten: Im ersten Schritt (Transition Interpret_Action) wird dabei die Aktion, die in der Request-Liste an erster Stelle steht, entsprechend der Funktion manual_request in ihre Bewegungsbestandteile „zerlegt" und der eigentliche motorische Prozess gestartet. Voraussetzung hierfür ist, dass auf dem Platz Hand_Free eine Marke liegt, d.h. zu diesem

[59] In der vorliegenden Modellversion sind keine Tirgger des motorischen Moduls aus anderen Modulen vorhanden, da das Aufnehmen von Flugstreifen aus dem Drucker und aus der Stripbay im durchgeführten Validierungsexperiment nicht Teil der Aufgabe der Versuchspersonen war. Die Version 1.8, die auf einem fiktiven Towerarbeitsplatz beruht, enthält allerdings diese Trigger.

Zeitpunkt also keine weiteren motorischen Prozesse laufen. In der Funktion `manual_request` ist für jede auftretende Aktion festgehalten, welche Locations für diese Aktion angesteuert werden. Auf den nachfolgenden Platz `Manual_Request` wird damit eine Liste abgelegt, in der alle Locations, die im Rahmen der Aktion angesteuert werden, aufgeführt sind.

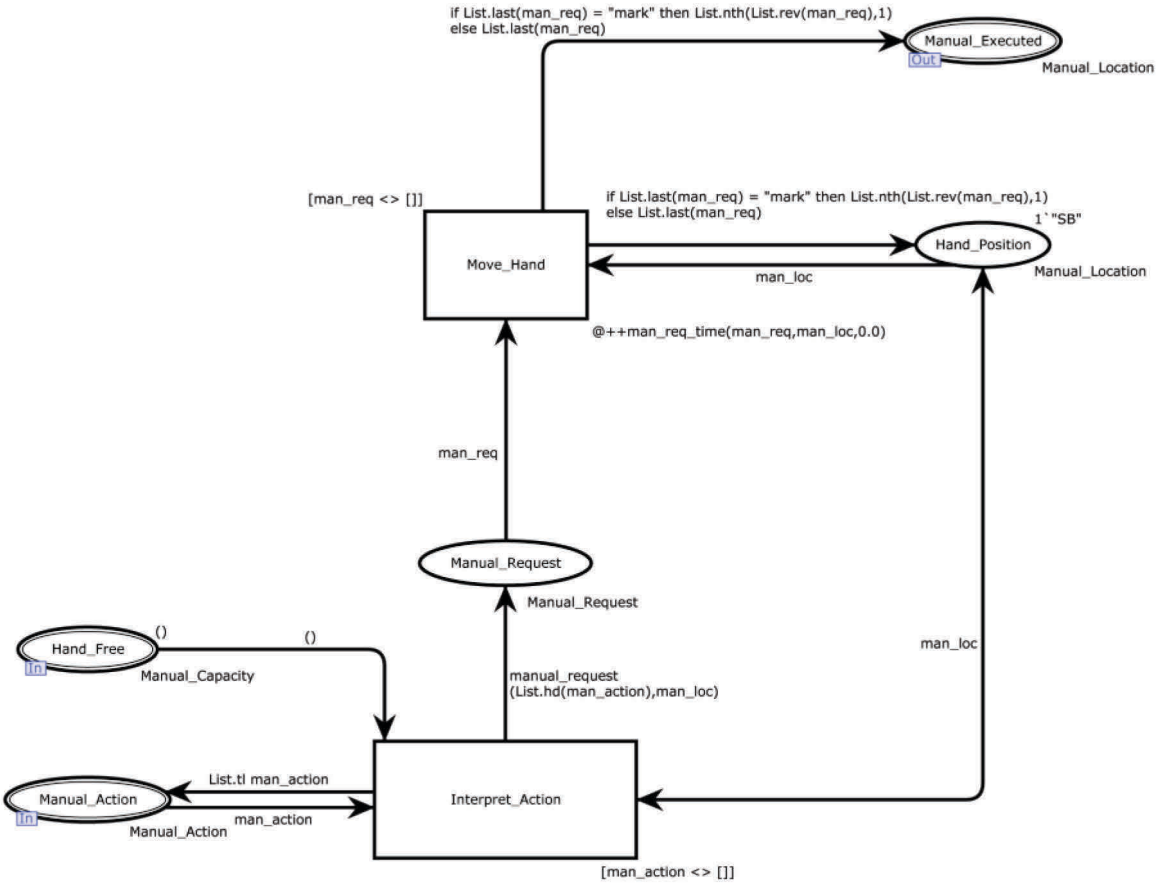

Abbildung 36: Das Manual-Modul im Petrinetz

Die folgende Transition `Move_Hand` gibt die Position der Hand nach Abschluss der gesamten Operation aus. Zusätzlich wird an dieser Stelle die Gesamtdauer der Operation berechnet. Diese Berechnung erfolgt anhand der Funktion `man_req_time`. Wie bereits erwähnt, dient als Input eine Liste aller Locations, die im Laufe der Operation angesteuert werden. Hierbei liegt die Annahme zugrunde, dass es sich um sehr gut geübte Bewegungen handelt, sodass alle Teilbewegungen in einer Gesamtbewegung zusammengefasst werden **(Annahme 19)**. Bei unbekannten bzw. wenig geübten Bewegungen wäre davon auszugehen, dass Teilbewegungen einzeln repräsentiert werden. Die rekursive Funktion `man_req_time` ermittelt die Dauer jeder daraus resultierenden Einzelbewegung und kumuliert diese. Die Dauer jeder möglichen Einzelbewegung ist entsprechend Kapitel 5.4.8 berechnet und in der Funktion `manual_time` aufgeführt.

Eine Ausnahme stellt hierbei die Bewegung „`mark`" dar, die das Markieren von Flightstrips oder einen Eintrag in den RSME-Fragebogen darstellt. Diese lässt sich nicht auf Grundlage von

Fitts' Law berechnen, da es sich nicht um eine klassische, zielende Bewegung handelt. Aus diesem Grund wurde die Dauer für einen `mark`-Vorgang auf 2s festgelegt **(Annahme 20)**. Die Hand selbst wird beim Feuern der Transition `Move_Hand` noch nicht freigegeben (keine Marke auf `Hand_Free`). Die Freigabe dieser Kapazität erfolgt auf der übergeordneten `act`-Ebene durch die jeweiligen Transitionen, welche die motorischen Bewegungen beschreiben (z.B. `Send_Clearance`) sowie Addition der Zeit, die durch das Manual-Modul berechnet wurde.

Sind die motorischen Bewegungen ausgeführt, so werden auf den jeweiligen nachgeschalteten Plätzen im `act`-Modul Marken generiert, welche die nachfolgenden Prozesse initiieren. Die Aktivierung des Funkgerätes initiiert damit, dass die Transition `Send_Clearance` feuern kann und damit eine Marke auf dem Platz `Radio_TWR` generiert wird, welche an das Interaktions-Modell übertragen wird und die Freigabe so an das Flughafenmodell übermittelt (vgl. Kapitel 5.3). Zusätzlich wird in Abhängigkeit der gegebenen Freigabe ein erneutes Überprüfen der freigegebenen Location, sowie das Abrufen der Information über den Zustand dieser Location (free oder occupied) initiiert (Plätze `check_Loc_status` und `Retrieve_Loc_status`). Zusätzlich wird über den Platz `Frequency_occupied` die Verwendung des Funkgerätes für eine gewisse Zeit geblockt (im Modell mit 5s festgelegt; mittlere Dauer einer Anfrage/Freigabe; vgl. Kapitel 5.3; **Annahme 21**). Außerdem wird wiederum in Abhängigkeit der zu erteilenden Freigabe Arbeitsgedächtniskapazität freigegeben (vgl. Annahme 14) .

Durch die Generierung einer Marke auf dem Platz `Clearance_delivered` wird das nachfolgende Markieren des Flugstreifens ermöglicht (Transition `Mark_Flightstrip`). Diese greift hierbei, ebenso wie die Transition `Fill_In_Questionaire`, auf die visuelle Aufmerksamkeit zu, da für die Bearbeitung dieser Aufgabe visuelle Ressourcen benötigt werden. Diese wird nach Abschluss der Flugstreifenmarkierung (Platz `Flightstrip_marked`) durch die nachgeschaltete Transition `Add_FS_to_SB_or_handover` wieder freigegeben. Zusätzlich wird hierbei je nachdem, ob weitere Aktionen des Luftfahrzeuges bearbeitet werden müssen, bestimmt, ob ein Handover stattfindet (Platz `Handover`) oder der Flugstreifen wieder in die Stripbay einsortiert wird (Platz `new_FS_Info`). Die Notwendigkeit des Einsortierens eines Flugstreifens löst wiederum den `plan`-Prozess aus (vgl. Kapitel 5.4.4 und 5.4.5).

5.4.10 Das Arbeitsgedächtnismodell

Das Arbeitsgedächtnismodell stellt für die in dieser Arbeit vorliegende Problemstellung die zentrale Komponente des Modells dar. Es wird dazu verwendet, eine quantitative Einschätzung der mentalen Beanspruchung des Lotsenmodells über die Zeit bereitzustellen. Das Modell selbst orientiert sich an den in Kapitel 2.2 vorgestellten Theorien des Arbeitsgedächtnisses. Aufgrund der zahlreichen ungeklärten Mechanismen und offenen Fragestellungen, vor allem zur

Ressourcenbegrenzung der einzelnen Teilsysteme, die im Zusammenhang mit Baddeleys Multi-Komponenten-Theorie (Baddeley & Hitch, 1974; Baddeley, 2012) stehen, wurde hier allerdings von einer einfachen Single-Limited Capacity Theory ausgegangen (vgl. Kapitel 2.2 und 3.4; **Annahme 22**). Es werden demnach keine unabhängigen Systeme für auditiv-verbale und räumlich-visuelle Informationen angenommen, sondern lediglich ein einzelner Speicher mit einer einzelnen begrenzten Kapazität (vgl. Kahnemann, 1973; Miller, 1956). Die Speicherung der Informationen erfolgt direkt nach der Informationsaufnahme aus dem `auditory` und `visual buffer` im `WM_storage`-Prozess (vgl. Abbildung 37).

Es wurde außerdem entscheiden, keine Modellierung eines wie bei Cowan (1999) beschriebenen zeitlichen „decay" der Informationen (Chunks) vorzunehmen, da seine Umsetzung im ereignisdiskreten Petrinetz einige Schwierigkeiten mit sich bringt: Einerseits müsste festgelegt werden, in welchen Zeitabständen ein in jedem Chunk enthaltener Aktivierungs-Wert diskret verringert wird, und zwar so lange, bis der Chunk vollständig verblasst (also vergessen) ist. Die Umsetzung eines solchen decay, selbst wenn er im Sekundenbereich läge, würde zu erheblichen Performanzeinbrüchen der Simulation führen.

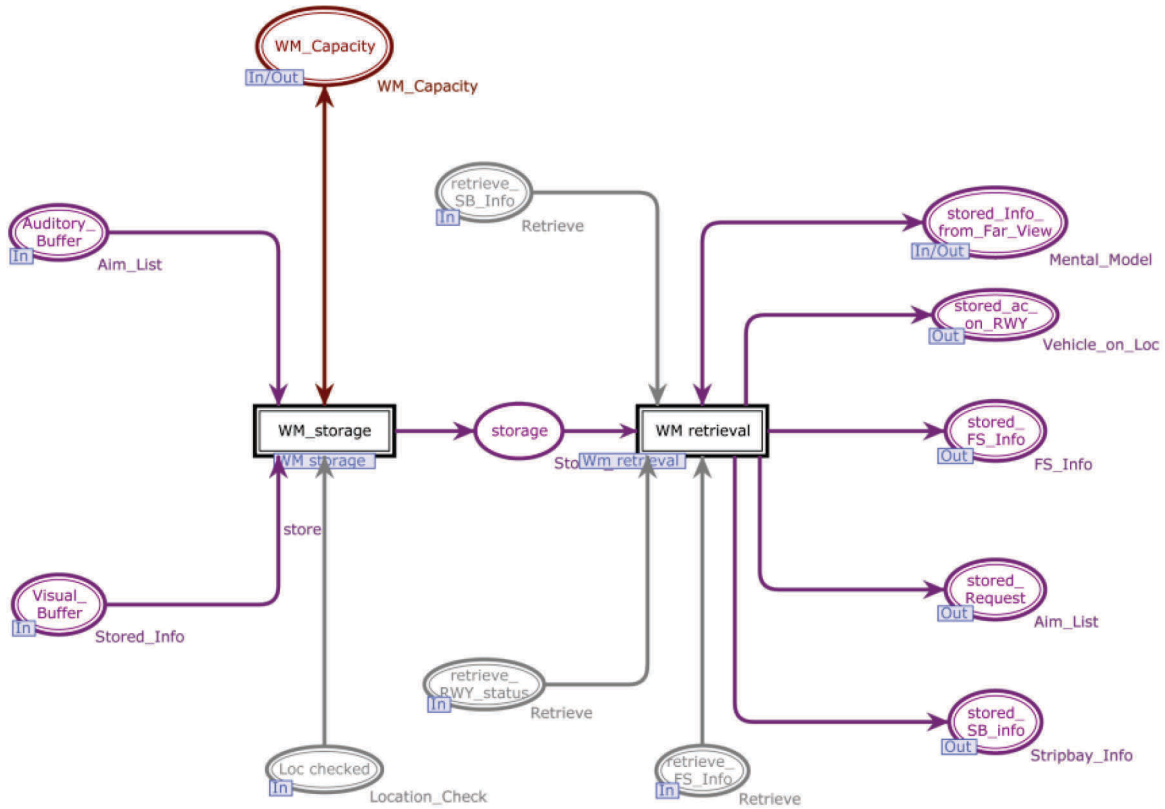

Abbildung 37: Das Working-Memory-Modul (WM) im Petrinetz

Weiterhin besteht das Problem, dass die im Arbeitsgedächtnis vorhandenen Chunks Trigger für nachfolgende Prozesse des kognitiven Modells darstellen (Marken als Vorbedingungen für das Schalten bestimmter Transitionen). Würde ein Chunk daher vollständig vergessen werden (Verschwinden der Marke), würde das Petrinetz in ein Deadlock laufen und das Netz damit tot sein. Es wird daher davon ausgegangen, dass die Informationen (Chunks) so lange im Speicher

liegen bleiben, bis alle kognitiven Prozesse, die in irgendeiner Weise die in ihnen enthaltenen Informationen benötigen, abgeschlossen sind. Erst im Anschluss daran können die Chunks „vergessen“ werden, was mit einer Freigabe der Arbeitsgedächtniskapazität einhergeht (`WM_Capacity`).

Der Abruf der Chunks (`WM_retrieval`) erfolgt erst dann, wenn die Information tatsächlich benötigt wird. Die Beschreibung eines solchen Abrufprozesses kann entfernt mit der bei Cowan (2001) beschriebenen Verschiebung eines Focus of Attention auf die jeweils benötigte Information verglichen werden. Hierbei wird allerdings nicht explizit davon ausgegangen, dass es einen einzelnen Focus of Attention gibt, der eine begrenzte Kapazität besitzt (vgl. auch McElree, 2001; Oberauer & Bialkova, 2009). Vielmehr wird lediglich der Prozess des Abrufens beschrieben, unter der Bedingung, dass vorgeschaltete Prozesse, die diese Information benötigen, bereits abgelaufen sind. Die genauen Abläufe und der Aufbau des Arbeitsgedächtnismoduls wird im folgenden Abschnitt beschrieben.

5.4.11 Arbeitsgedächtnis im Petrinetzmodell

Das Petrinetzmodell des Arbeitsgedächtnisses (`WM`) besteht aus 13 Plätzen und zwei Transitionen und ist nochmals in zwei Sub-Module, die die Speicherung (`WM_storage`) und den Abruf (`WM_retrieval`) der Chunks beschreiben, unterteilt. Diese bestehen jeweils nochmals aus sechs Pläzen und Transitionen (`WM_storage`) bzw. neun Plätzen und fünf Transitionen (`WM_retrieval`). Die in der Phase der Informationsaufnahme erhaltenen Informationen werden in Form von Marken jeweils auf den Eingangsplätzen `Auditory_Buffer` und `Visual_Buffer` abgelegt und dem `WM_storage`-Modul zur Verfügung gestellt (siehe Abbildung 38 auf Seite 128). Hierbei werden die auf den Buffer-Plätzen abgelegten Informationen durch das Schalten der jeweiligen Transition (z.B. `store_aud` für die Speicherung auditiver Informationen, also der Anfrage) in den Speicher (`Storage`) verschoben. Dabei wird für jede Information (Chunk) auf dem Platz `WM_Capacity` die Anzahl der hier gespeicherten Elemente erhöht. Dies wird durch eine Marke vom colorset `int` realisiert, welcher bei Initialisierung des Netzes ein Wert von „0“ zugewiesen wird. Mit jedem gespeicherten Chunk (Auftreten einer `store`-Transition) wird dieser Wert um 1 erhöht. Ein Chunk wird dabei mit einer Marke gleichgesetzt, welche weitere Items enthält. Hierbei können unterschiedliche Marken unterschiedlich komplexe Chunks enthalten (siehe Beschreibung der Colorsets Kapitel 5.2.3). Beispielsweise enthält die Marke vom Typ `Aim_List` auf dem Platz `auditory_buffer` die Informationen, die einer Anfrage entsprechen: Anfragendes Flugzeug (Rufzeichen), letzte Position dieses Flugzeuges, Zielposition des Flugzeuges, Intention und die Zeit, zu der die Anfrage gestellt wurde.

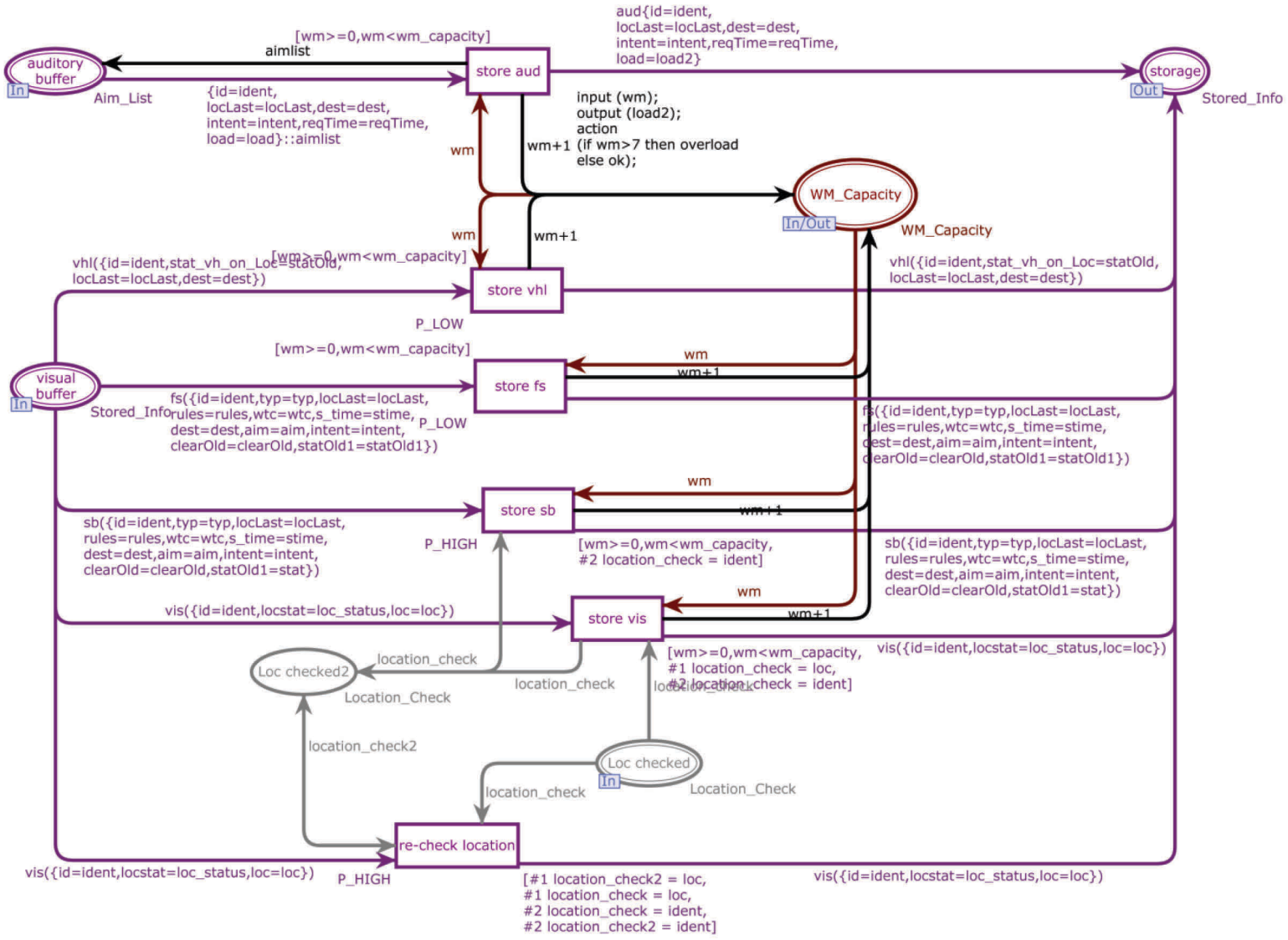

Abbildung 38: Das WM_Storage-Modul im Petrinetz

Der Platz `visual_buffer` kann einerseits Informationen von den Flugstreifen, andererseits Informationen aus der Außensicht enthalten. Informationen von Flugstreifen sind sehr komplex - sie enthalten im Modell die Informationen zu einem Flugzeug:

- Rufzeichen
- Flugzeugtyp (z.B. A319)
- letzte Position
- nach welchen Regeln das Flugzeug unterwegs ist (IFR oder VFR)
- wie die Wirbelschleppenkategorie des Flugzeuges ist (wtc)
- die geplante Ankunfts- bzw. Abflugszeit
- Zielposition
- Ziel (z.B. Arrival oder Departure)
- Intention (z.B. Taxi_to_RWY oder Taxi_on_Block)
- bisherige vorliegende Freigabe
- letzter Status (z.B. cleared_to_land)

Die Übernahme der kompletten Colorsets als Chunks beruht dabei auf der Annahme, dass Lotsen alle Informationen über ein Flugzeug in einem Chunk abspeichern **(Annahme 23)**. In der Realität ist allerdings davon auszugehen, dass der Lotse nur diejenigen Informationen eines Chunks verwendet, die er tatsächlich für die Bearbeitung seiner Aufgabe benötigt (z.B. nur die geplante Ankunftszeit bei sortieren der Flugstreifen oder nur das Rufzeichen, sowie die Intention, die letzte Position und die beabsichtigte Zielposition bei Landungen). Diese Vorgehensweise hat zusätzlich auch praktische Gründe. Bei Veränderung der Colorsets (z.B. Verringerung der Informationsanzahl innerhalb des Colorsets) würden bei der Übergabe an das Interaktionsmodell Informationen fehlen. Im Sinne einer reibungslosen Kommunikation der einzelnen Modellteile untereinander ist es daher sinnvoll, die vollständigen Colorsets zu belassen und alle Informationen durch das gesamte Lotsenmodell weiterzureichen.

Das gleiche Problem besteht auch, würde man individuelle Unterschiede in der Fähigkeit Chunks zu bilden abbilden wollen. Hierbei wäre auch eine Veränderung der Colorsets nötig, indem Informationen begrenzt werden. Da festgestellt wurde, dass nicht nur die Anzahl der Items innerhalb eines Chunks, sondern auch die maximale Anzahl der Chunks mit steigender Expertise steigt (vgl. Kapitel 2.2), kann über die flexibel anpassbare Arbeitsgedächtniskapazität zumindest die maximale Anzahl speicherbarer Chunks verändert werden. Die Konstante `val wm_capacity`, im CPN-Tools-Programmierbereich unter `Controller` → `Variables` zu finden, kann hierbei auf beliebige Werte festgelegt werden. Es sollte allerdings psychologische Plausibilität gewährleistet sein. Für den zur Modellvalidierung durchgeführten Versuch wurde

die Kapazität auf zehn Chunks festgelegt, da zahlreiche Befunde mit Lotsen von einer maximalen Kapazität von zehn Chunks ausgehen (Bainbridge, 1975; Bisseret, 1971; Sperandio, 1969).

Zusätzlich zu den flugsicherungsspezifischen Informationen enthält das zusammengesetzte Colorset `Aim_List` das einfache Colorset `load`. Dieses dient der Selbstregulation des Modells. Wie Sperandio (1978) beschreibt, und wie aus der Handlungsregulationstheorie (Hacker, 1986) hervorgeht, verfügen Menschen über spezielle Regulationsmechanismen, welche sie dazu befähigen, die richtigen Handlungen auszuführen, um ein antizipiertes Ziel zu erreichen (vgl. auch Desmond & Hoyes, 1996; Loft et al., 2007).

Einen solchen Regulationsmechanismus beschreibt Sperandio (1978), der in einer Untersuchung herausfand, dass Lotsen in Abhängigkeit ihrer mentalen Beanspruchung ihre Handlungsstrategien anpassen. Die mentale Beanspruchung dient hierbei als unabhängige Variable, welche zu einer Anpassung der Arbeitsmethoden führt, um kognitive Überlastung zu verhindern und die Situation zeitlich nach hinten zu verschieben.

Ein solcher *low risk action regulation mechanism* (vgl. Smieszek et al., 2013) wurde in das Modell implementiert, um eine Überlastung des Modells und damit auch ein Deadlock des Petrinetzes zu verhindern. Das Modell wird dabei bei einer festgelegten Anzahl an Chunks (`wm_capacity - 3`) im Arbeitsgedächtnis keine weiteren Anfragen bearbeiten, sondern den anfragenden Flugzeugen lediglich mitteilen, dass sie in einigen Minuten erneut anfragen sollen. Für den durchgeführten Validierungsversuch spielt dieser Mechanismus allerdings keine Rolle, da die maximale kritische Anzahl Chunks im Modell nicht erreicht wird.

Sind die Chunks im Speicher (Platz `storage`) abgelegt, stehen sie dem nachfolgenden Abrufprozess (`WM_retrieval`; siehe Abbildung 39) zur Verfügung. Hier werden die im Speicher abgelegten Informationen über die jeweilige retrieval-Transition abgerufen, wenn sie für weitere Informationsverarbeitungsprozesse (z.B. Entscheidungen) benötigt werden. Hierbei wird davon ausgegangen, dass Bottom-Up Informationen (Anfrage eines Flugzeuges; Transition `retrieve_req`, sowie Feststellen eines Flugzeuges auf der Start-/Landebahn, das eine Folgefreigabe benötigt; Transition `retrieve_vhl_on_loc`) ohne weitere Voraussetzung sofort abgerufen werden. Alle anderen Informationen benötigen Top-Down Trigger (Transitionen und gleichnamige Plätze `retrieve_RWY_status`; `retrieve_FS_info`; `retrieve_SB_info`). Ist der retrieval-Prozess abgeschlossen, stehen die Informationen für die nachfolgenden Verarbeitungsschritte (`plan`; siehe Kapitel 5.4.4; `decide`, siehe Kapitel 5.4.6) zur Verfügung.

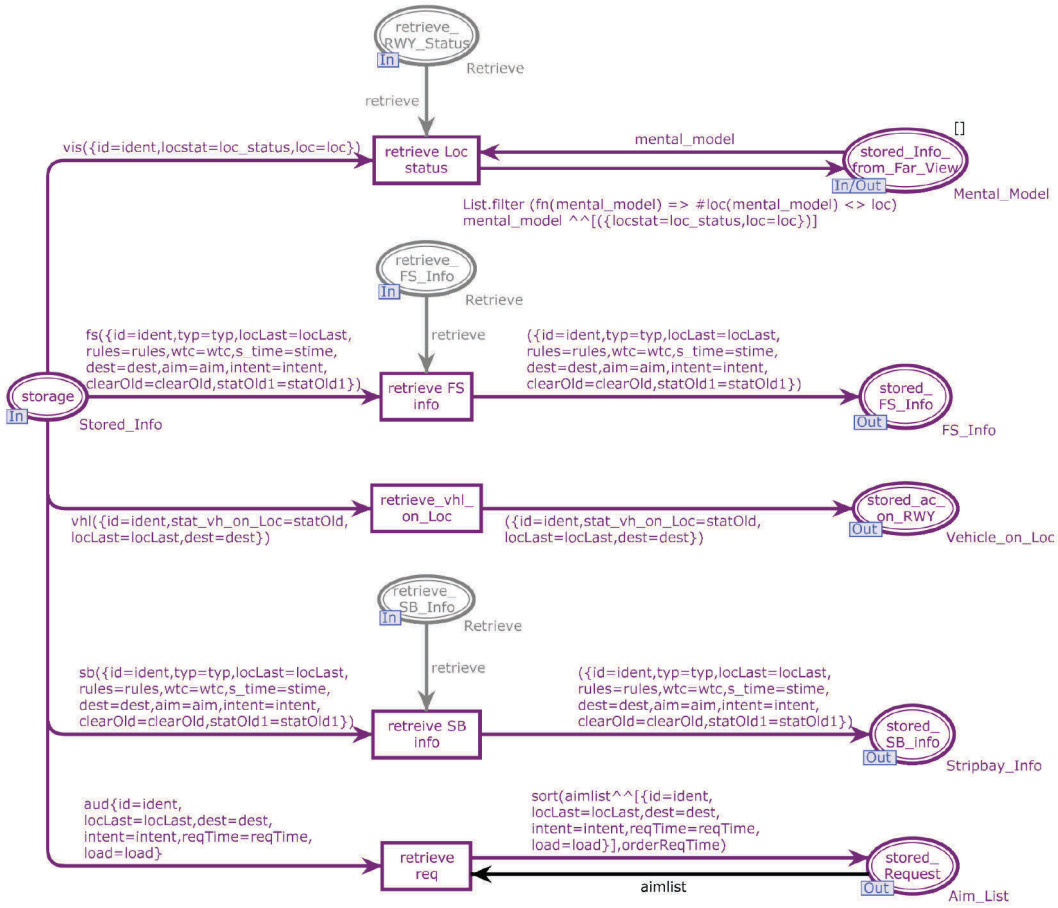

Abbildung 39: Der WM_retrieval-Prozess im Petrinetz

5.5 Simulation und Analyse des Petrinetzmodells

```
1  0.0 WM_Capacity 1`0
2  1.0 WM_Capacity 1`0
3  1.0 WM_Capacity 1`0
4  6.0 WM_Capacity 1`0
5  6.0 WM_Capacity 1`0
6  6.0 WM_Capacity 1`0
7  6.0 WM_Capacity 1`0
8  6.0 WM_Capacity 1`1
9  6.0 WM_Capacity 1`1
10 6.0 WM_Capacity 1`1
11 6.0 WM_Capacity 1`1
12 6.0 WM_Capacity 1`1
13 9.9 WM_Capacity 1`1
14 9.9 WM_Capacity 1`1
15 9.9 WM_Capacity 1`2
16 9.9 WM_Capacity 1`2
17 9.9 WM_Capacity 1`1
18 17.1 WM_Capacity 1`1
19 17.1 WM_Capacity 1`1
```

Abbildung 40: Ausgabe des WM_ Capacity-Monitors in einer Textdatei

Wie im vorherigen Kapitel erwähnt, erfolgt die Quantifizierung der mentalen Beanspruchung des Lotsenmodells anhand der Anzahl der Chunks im Arbeitsgedächtnis. Bei der Simulation des Petrinetzmodells erfolgt dabei die Ausgabe von Logdateien in Form von *.txt files. Um diese Dateien während der Simulation online erstellen zu können, bietet CPN-Tools die Möglichkeit sogenannte Monitore[60] an die gewünschten Plätze und Transitionen anzuhängen. Diese ermöglichen es Daten aus dem Modell auszulesen und sie in *.txt-files zu speichern. Für die Ausgabe der Arbeitsgedächtniskapazität wurde ein einfacher *write-in-file Monitor* am Platz `WM_Capacity` definiert, welcher für jeden Modellschritt die jeweilige Modellzeit sowie die momentane Arbeitsgedächtnisbelegung (als Markenbelegung) ausgibt (siehe Abbildung 40).

[60] Für eine genauere Beschreibung der Monitore sei hier auf Jensen und Kristensen (2009), sowie die CPN-Tools Homepage www.cpntools.org verwiesen.

Darüber hinaus wurden noch zwei weitere Monitore an den Transitionen `Requesting_TWR` im Interaktions-Modell, sowie `Send_Clearance` im `act`-Modul des Lotsenmodells definiert. Diese schreiben jeweils beim Schalten der Transition den Inhalt der jeweiligen transportierten Marke in eine Textdatei (vgl. Abbildung 41).

```
1.0 Request ("COA1598",1,Taxi_to_RWY,Flightstrip_printed,(Stand,U,14,Both),(RWY,R,19,Std),[])
26.65 Request ("EGF3172",22,Taxi_to_RWY,Flightstrip_printed,(Stand,B,14,Both),(RWY,R,19,Std),[])
99.0 Request ("EGF3189",99,Landing,Flightstrip_printed,(Sky,R,19,Std),(Stand,A,1,Both),[])
132.2 Request ("EGF3172",22,Departure,Flightstrip_printed,(Holdingpoint,L,19,Std),(RWY,R,19,Std),[])
177.0 Request ("AAH485",177,Landing,Flightstrip_printed,(Sky,R,19,Std),(Stand,A,3,Both),[])
190.65 Request ("COA1598",1,Departure,Flightstrip_printed,(Holdingpoint,L,19,Std),(RWY,R,19,Std),[])
256.0 Request ("AAH481",256,Landing,Flightstrip_printed,(Sky,R,19,Std),(Stand,A,4,Both),[])
305.0 Request ("DAL1778",305,Taxi_to_RWY,Flightstrip_printed,(Stand,U,15,Both),(RWY,R,19,Std),[])
335.0 Request ("SWA148",335,Landing,Flightstrip_printed,(Sky,R,19,Std),(Stand,A,2,Both),[])
345.0 Request ("AWE3",343,Taxi_to_RWY,Flightstrip_printed,(Stand,B,10,Both),(RWY,R,19,Std),[])
411.8 Request ("SWA212",409,Landing,Flightstrip_printed,(Sky,R,19,Std),(Stand,B,9,Both),[])
435.55 Request ("AWE3",343,Departure,Flightstrip_printed,(Holdingpoint,L,19,Std),(RWY,R,19,Std),[])
509.0 Request ("DAL1778",305,Departure,Flightstrip_printed,(Holdingpoint,L,19,Std),(RWY,R,19,Std),[])
537.0 Request ("AWE652",537,Landing,Flightstrip_printed,(Sky,R,19,Std),(Stand,B,12,Both),[])
582.0 Request ("AAL2193",582,Taxi_to_RWY,Flightstrip_printed,(Stand,C,21,Both),(RWY,R,19,Std),[])
626.0 Request ("EGF3165",626,Landing,Flightstrip_printed,(Sky,R,19,Std),(Stand,C,21,Both),[])
```

Abbildung 41: Ausgabe des Requesting_TWR-Monitors in einer Textdatei

Diese Dateien geben Aufschluss darüber, wann eine Anfrage aus dem Flughafenmodell das Lotsenmodell erreicht hat und wann das Lotsenmodell mit der entsprechenden Freigabe geantwortet hat. Sie können demnach als eine Art Funkprotokoll des MATriCS-Modells angesehen werden.

5.6 Kapitelzusammenfassung

Die Entwicklung des MATriCS-Modells erfolgte unter Berücksichtigung der zwölf Schritte zur Erstellung von CPN-Modellen nach Jensen (1997a) und ist mit der Erstellung eines Computerprogramms vergleichbar. Es erfolgte zunächst eine Dreiteilung des Modells auf oberster Ebene in Flughafen-, Interaktions- und Lotsenmodell in Anlehnung an Cacciabue (1998b) und Werther (2006a).

Das *Flughafenmodell* wurde in Zusammenarbeit mit dem Deutschen Zentrum für Luft- und Raumfahrt in Braunschweig als generisches Modell entwickelt und ist in der Lage nahezu jeden beliebigen Flughafen und jedes beliebige Verkehrsszenario zu beschreiben (vgl. Kapitel 5.2). Der Austausch von Informationen zwischen dem Flughafenmodell und dem Lotsenmodell erfolgt über das *Interaktionsmodell*, in dem die wichtigsten Informationsquellen für den Lotsen (Funkkommunikation, Außensicht, Radar und Flugstreifen) modellhaft abgebildet sind (vgl. Kapitel 5.3).

Das *Lotsenmodell* stellt die eigentliche kognitive Simulation dar und bildet daher die kognitiven Prozesse des Lotsen ab (vgl. Kapitel 5.4). Zusätzlich werden auch kognitive Ressourcen modelliert, welche für den vorliegenden Anwendungsfall speziell von Interesse sind. Grundlegend orientiert sich das Lotsenmodell an der Handlungsregulationstheorie (Hacker, 1986), sowie an der in Kapitel 4.3 durchgeführten Aufgabenanalyse und dem in Kapitel 4.5 beschriebenen theoretischen Modell. Dabei wurde das Lotsenmodell nicht exakt in dem

Detailgrad und der Tiefe umgesetzt, wie diese in der Aufgabenanalyse und dem theoretischen Modell beschrieben wurden, sondern nur so detailliert wie nötig.

Die einzelnen Phasen der Handlungsregulation wurden im Modell als Sub-Module umgesetzt. Eingehende Informationen werden über das Perceive-Modul aufgenommen, wobei hierbei begrenzte Aufmerksamkeitsressourcen für visuelle und auditive Indformationsaufnahme angenommen werden (vgl. Wickens, 1984). Eine Verschiebung der visuellen Aufmerksamkeit erfolgt anhand festgelegter Prioritäten nach dem SEEV-Modell (Wickens & McCarley, 2008; Wickens et al., 2007, 2001) hin zu verschiedenen Areas of Interest (Außensicht, Radar, Flugstreifen). Priorität liegt dabei auf der Außensicht und dem Radar (vgl. Kapitel 5.4.2 und 5.4.3). Die so erhaltenen Informationen werden im nachfolgenden Working-Memory-Modul gespeichert. Die gespeicherten Informationen stehen den nachfolgenden Planungs- und Entscheidungsprozessen zur Verfügung. Die Planung beschreibt das Sortieren der Flugstreifen anhand der geplanten Ankunftszeit (vgl. Kapitel 5.4.4 und 5.4.5). Für die Entscheidungsfindung wurde eine einfache „first come, firtst served"-Heuristik implementiert (vgl. Möhlenbrink, 2011), wobei in Entscheidungen für Starts und Landungen, sowie das Rollen unterschieden wird (vgl. Kapitel 5.4.6 und 5.4.7). Die getroffene Entscheidung wird in der Phase der Handlungsausführung durch Übermittlung der Freigabe an das Flughafenmodell weitergegeben. Hierbei sind physische Bewegungen notwendig (z.B. zum Aktivieren des Funkgerätes oder Markieren der Flugstreifen), die ebenfalls im motorischen Modul abgebildet sind (vgl. Kapitel 5.4.8 und 5.4.9). Die Speicherung und zentrale Verarbeitung der Informationen erfolgt im Sub-Modul Working Memory, welches die Speicherung und den Abruf von Informationseinheiten beschreibt und eine begrenzte Kapazität besitzt. Letzteres ist für die Quantifizierung der mentalen Beanspruchung zentral, weshalb das Auslesen der Arbeitsgedächtnisbelastung während der Simulation in diesem Sub-Modul erfolgt (vgl. Kapitel 5.4.10 und 5.4.11). Ferner können außerdem Funkdaten aus dem Modell ausgelesen werden, welche die Zeitpunkte des Auftretens von Anfragen und Freigaben protokollieren. Mit der Implementierung des Modells wurde der sechste Schritt des Cacciabue-Frameworks bearbeitet. Der in Kapitel 4.1 hinzugefügte siebte Schritt der Modellvalidierung wird im folgenden Kapitel beschrieben.

6 VALIDIERUNG

Wie bereits in Kapitel 4.1 beschrieben besteht die Aufgabe des Modellierers nicht nur darin, Modelle zu erstellen, sondern ebenso über die Validierung den Bezug zwischen dem Modell und dem realen Untersuchungsgegenstand nachzuweisen (Bub & Lugner, 1992; Möhlenbrink, 2011). Hierbei sind durch empirische Studien Daten zu generieren, welche mit den Modelldaten verglichen werden können, um Informationen über die Modellgüte zu erhalten.

Eine solche Studie wurde im August 2013 in den Räumlichkeiten des Institutes für Psychologie und Arbeitswissenschaften der TU Berlin durchgeführt. Ziel der Studie war die Erhebung von Daten zur Testung des Modellansatzes und der Modellgüte, sowie die Replikation von Befunden mentaler Beanspruchung in Abhängigkeit des Verkehrsaufkommens. Bei der Studie ging es darum zu testen, wie gut das entwickelte Modell die bei den Versuchspersonen erhobenen Daten zur mentalen Beanspruchung bei einer Lotsenaufgabe abbildet. Daher wurden verschiedene Beanspruchungsmaße erhoben (vgl. Kapitel 2.4). Als Beanspruchungsmaße wurden sowohl unidimensionale als auch multidimensionale subjektive Fragebogendaten erhoben, sowie als objektives Maß das EKG gemessen, woraus als Beanspruchungsmaß die Herzrate, gemessen in Inter-Beat-Intervallen, abgeleitet wurde (siehe Kapitel 2.4 zur Erfassung mentaler Beanspruchung). Als Leistungsmaß wurde die Anzahl der Fehler gemessen. Zur Quantifizierung der Modellgüte wurden die so erhobenen Maße mentaler Beanspruchung mit den Modelldaten verglichen und Goodness-Of-Fit-Maße berechnet. Zur Überprüfung der experimentellen Manipulation (Replikation der Befunde zu mentaler Beanspruchung in Abhängigkeit des Verkehrsaufkommens) wurden statistische Tests durchgeführt.

Im Methodenkapitel 6.1 werden zunächst die Teilnehmercharakteristika (6.1.1), anschließend der Simulator und das Material (6.1.2), das Versuchsdesign (6.1.3), die Hypothesen (6.1.4) und der Versuchsablauf (6.1.5) beschrieben. Es folgt die Darstellung der Ergebnisse geordnet nach Hypothesen (Kapitel 6.2), sowie im Anschluss die Interpretation der gewonnenen Ergebnisse entsprechend der Hypothesen (Kapitel 6.3). Zum Schluss wird die Studie im Kapitel 6.4 zusammengefasst.

6.1 METHODIK

Um die Beanspruchungsdaten zu erheben, hatten Versuchsteilnehmer in einer Towersimulation vier verschiedene 15-minütige Verkehrsszenarien zu bearbeiten. Ihre Aufgabe bestand darin, den Verkehr, also Landungen und Starts, sowie das Rollen am Boden, an einem kleinen Flughafen zu regulieren. Hierbei war es vor allem wichtig, die Aufgabe für die Versuchsteilnehmer so realistisch wie möglich abzubilden, dass also die verwendete Towersimulation ein möglichst realistisches Abbild des Arbeitsplatzes von Towerlotsen

ermöglicht, wobei gleichzeitig gewährleistet sein musste, dass die Aufgaben in der Studie den im Modell abgebildeten Aufgaben gleichen. Nur wenn dies gewährleistet ist, ist es möglich einen fairen Vergleich der Beanspruchungsdaten von Modell und Lotse anzustellen.

6.1.1 TEILNEHMER/STICHPROBE

Da es sich als äußerst schwierig erwies eine größere, für die Studie benötigte, Anzahl Towerlotsen für den Versuch zu rekrutieren und die finanziellen Mittel begrenzt waren, wurde der Versuch anhand einer Laien-Stichprobe durchgeführt. Dabei wurde gewährleistet, dass die Versuchspersonen durch ausführliche Instruktion und Training möglichst viel Routine und Übung erhielten. Die Verwendung einer Laienstichprobe erweist sich außerdem dahingehend als sinnvoll, als dass nicht alle Regeln der Flugverkehrskontrolle bei der Versuchsdurchführung adressiert werden mussten, da auch im Modell nur vereinfachte Regeln umgesetzt wurden.

Der größte Teil der Teilnehmer für die Studie wurden über den Probandenserver des Graduiertenkollegs prometei gewonnen. Ein geringer Teil wurde über schriftliche Ausschreibungen an den Instituten der TU Berlin gewonnen. Die Stichprobengröße wurde anhand der von Bortz und Döring (2006, S. 615) beschriebenen Tabelle anhand einer einfaktoriellen Varianzanalyse festgelegt. Hierbei empfehlen Bortz und Döring (2006) für eine einfaktorielle Varianzanalyse mit Messwiederholung mit drei Freiheitsgraden und einem Signifikanzniveau von $\alpha = 0{,}05$, einer durchschnittlichen Korrelation von $\bar{r} = 0{,}5$ ($1-\beta = 0{,}8$) und einem mittleren zu erwartenden Effekt eine Stichprobengröße von 23 Versuchspersonen. Um die vollständige Permutation aller Szenarien zu gewährleisten (vgl. Kapitel 6.1.3) wurde die Anzahl der benötigten Versuchspersonen auf 24 festgelegt.

Insgesamt nahmen 24 Versuchspersonen am Experiment teil. Hiervon waren 11 weiblich und 13 männlich mit einem durchschnittlichen Alter von 26,95 Jahren (Spannweite von 22 – 39 Jahren; $s = 4{,}69$). Der größte Teil der Teilnehmer waren Studenten (17), weitere fünf Personen waren Berufstätig, eine Person war Schüler, ein weiterer machte keine Angabe. Weiterhin gaben fast alle (23) der Versuchspersonen an, keinerlei Erfahrung mit Flug- oder Towersimulationen zu besitzen. Nur eine Versuchsperson gab an viel Erfahrung mit Flugsimulatoren zu besitzen. 14 Versuchspersonen verwendeten keine Sehhilfe, neun benötigten eine Sehhilfe und verwendeten diese auch. Eine Versuchsperson machte keine Angabe hierzu. Weiterhin waren 20 Versuchspersonen Rechtshänder, zwei Linkshänder, zwei machten hierzu keine Angabe.

6.1.2 SIMULATOR UND MATERIAL

Da kein High-Fidelity-Towersimulator[61] zur Verfügung stand, musste als Simulationssoftware auf Simulatoren für Privatanwender zurückgegriffen werden (Mid-Fidelity-Simulation). Röbig, König und Hofmann (2010) untersuchten einige solcher Simulatoren auf Ihre Nutzung zum Aufbau eines „low-cost Towersimulators". Hierbei werden unter anderem die Simulation Tower Simulator (Wilco Publishing, 2008), Microsoft Flight Simulator und ATC Airport Tower Simulator beschrieben.

Die Simulation hatte hierbei folgende Kriterien zu erfüllen:

- Möglichst realistisches Abbild der Aufgaben von Towerlotsen
- Hohe Übereinstimmung mit den im Modell abgebildeten Aufgaben
- Vorhandensein eines Flughafens mit einfachem Layout
- Einfache Erstellung und individuelle Anpassung von Verkehrsszenarien
- Vorhandensein einer Single-Player-Umgebung (keine Notwendigkeit von pseudo-Piloten und eines damit verbundenen komplexen Versuchsaufbaus)
- Direkte Kommunikation zwischen Versuchspersonen und Simulation über Spracherkennung
- Möglichkeit der Erzeugung einer hohen mentalen Beanspruchung

Die Simulation „Tower Simulator" wird bei Röbig, König und Hofmann (2010) als sehr realistisch beschrieben. Sie ermöglicht dem Anwender das Arbeiten als Towerlotse an fünf verschiedenen Flughäfen mit unterschiedlich komplexem Layout, wobei zwei Flughäfen mit wenig komplexem Layout vorhanden sind. Als Arbeitsmittel stehen Boden- und Luftradar, sowie eine Außensicht zur Verfügung, wobei die Darstellung letzterer unter heutigen Gesichtspunkten veraltet scheint. Die Anfragen von Piloten werden über ankommende Funksprüche und Textnachrichten dargeboten. Die Kommunikation des Anwenders mit der Simulation findet über eine Texteingabebox mit Hilfe standardisierter Befehle statt. Prinzipiell soll die Software die Anbindung einer Stimmerkennungssoftware unterstützen, was sich im vorliegenden Fall allerdings als technisch nicht möglich herausstellte. Ein großer Vorteil des Tower Simulator besteht darin, dass der Aufwand, eine Multiplayerumgebung aufzubauen, entfällt (die beispielsweise bei Microsoft Flight Simulator nötig ist). Zusätzlich ist die Erstellung individueller Flugpläne sehr leicht durch die Anpassung der vorgegebenen Default-Flugpläne der jeweiligen Flughäfen zu gewährleisten.

[61] Als High-Fidelity-Simulation wird eine Simulation angesehen, die es ermöglicht die Aufgaben und Arbeitsumgebung von Operateuren hochgradig realistisch dazustellen, im Gegensatz zu einer Mid-Fidelity-Simulation, bei der einige Vereinfachungen gemacht werden müssen.

Als problematisch stellte sich allerdings eine Eigenschaft des Tower Simulator heraus: Trotz der einfachen Erstellung von Flugplänen und der darin enthaltenen Angabe zum Zeitpunkt des Erscheinens der Flugzeuge (Setzen der erwarteten Ankunftszeit), konnten mit der Simulation keine standardisierten Szenarien erstellt werden. Zwar konnte in jedem Szenario für jedes Flugzeug eine erwartete Ankunftszeit angegeben werden, und somit ein einheitlicher Flugplan für alle Versuchspersonen zugrunde gelegt werden. Dieser galt für die Software allerdings nur als Richtwert, sodass das tatsächliche Auftauchen eines Flugzeuges bei laufender Simulation zufällig um diesen angegebenen Zeitpunkt variierte (± ca. 2 Minuten um den im Flugplan angegebenen Zeitpunkt)[62].

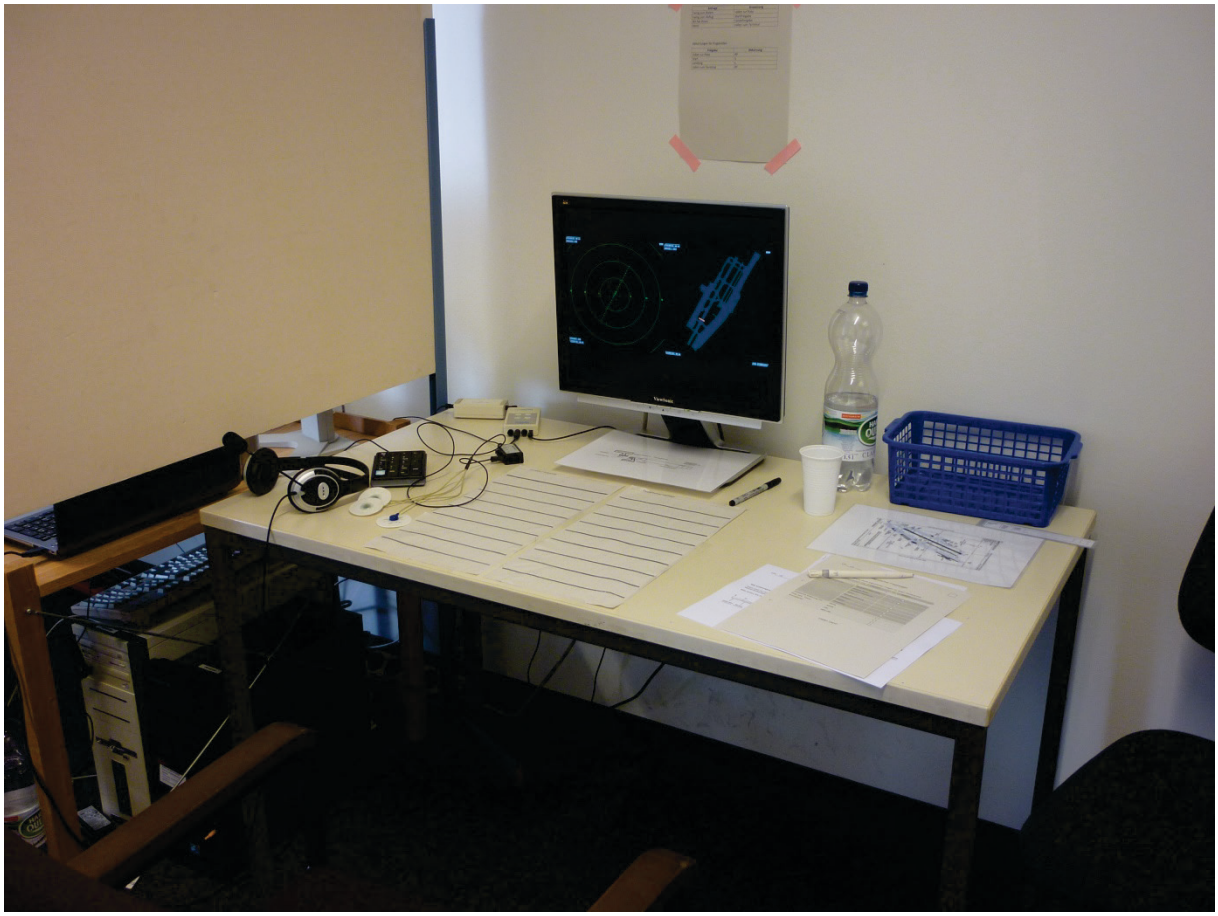

Abbildung 42: Arbeitsplatz der Versuchspersonen

Eine absolute Standardisierung der Szenarien wäre allerdings ohnehin nicht möglich gewesen, da auf Grund der Realitätsnähe der Aufgabe die Versuchspersonen selbst durch ihre Aktionen während der Szenarien Veränderungen an denselben vornehmen. Das bedeutet, dass je nachdem wie die Versuchsperson das Szenario abarbeitet und welche Entscheidungen sie trifft, das Szenario einen anderen Ablauf erhält. Dies stellt einen Trade-off zwischen der Kontrolle

[62] Eine Kontaktaufnahme mit dem Entwickler sowie dem Publisher der Software ergab leider keine Abhilfe für dieses Problem, da der Support schon seit längerem völlig eingestellt wurde und die Zuständigkeit von jeder Seite an die jeweils andere Seite verwiesen wurde. Es galt daher mit den Gegebenheiten bestmöglich umzugehen und die Auswertung der Daten dementsprechend zu gestalten (vgl. Kapitel 6.1.4).

aller Variablen und der Übereinstimmung mit der Realität dar, der zugunsten der Übereinstimmung mit der Realität entschieden wurde. Bei beiden Problemen erweist sich allerdings die Flexibilität des Flughafenmodells als äußerst hilfreich, da es möglich ist, aus den Funkprotokollen der Versuchspersonen exakt diejenigen Szenarien zu rekonstruieren, welche die konkrete Versuchsperson erlebt hat. Das heißt, es ist möglich, die Flugzeuge in den Modellszenarien exakt zu denselben Zeitpunkten erscheinen zu lassen, wie es bei den Versuchspersonen der Fall war. Diese können individuell in das Szenario eingelesen werden, so dass das Modell exakt das gleiche Szenario bearbeitet, wie die Versuchsperson.

Zusätzlich zu dem vom Simulator zur Verfügung gestellten Radarbildschirmen hatten die Versuchspersonen eine physische Stripbay mit Flugstreifen zur Verfügung, sowie ein Headset zur Funkkommunikation, wobei immer dann, wenn die Versuchspersonen sprechen wollten, eine blau markierte Taste auf einem Nummernpad gedrückt gehalten werden musste. Nicht mehr benötigte Flugstreifen (abgearbeitete Flugzeuge) konnten in einem Körbchen „entsorgt“ werden. Den ganzen Versuch über standen außerdem eine Karte des Flughafens, sowie zur Gedankenstütze eine Übersicht über mögliche auftretende Anfragen und dazugehörige Freigaben zur Verfügung. Da das Flughafenlayout ebenfalls einen Komplexitätsfaktor darstellt, der sich auf die mentale Beanspruchung der Versuchspersonen auswirken kann (vgl. Kapitel 2.3), wurde ein Flughafen mit geringer Komplexität ausgewählt, um eine Überforderung der Versuchspersonen durch zu viele Informationen und parallele Aufgaben (wie es zum Beispiel beim Kontrollieren von zwei Start-/Landebahnen der Fall wäre) zu vermeiden. Zusätzlich wird in der Realität die Kontrolle größerer Flughäfen (mit zwei Start-/Landebahnen) nie einem einzigen Lotsen überlassen. Die Kontrolle eines kleinen Flughafens kann jedoch, vor allem bei Regionalflughäfen, auch durch einen Lotsen erfolgen (vgl. Kapitel 4.3.6). Als Simulationsflughafen wurde daher der in der Software enthaltene Flughafen Santa-Ana, Kalifornien, USA (ICAO-Code: KSNA) verwendet, da dieser nur eine Start-/Landebahn und einen Hauptrollweg von und zu den Parkpositionen besitzt.

6.1.3 VERSUCHSDESIGN

Das Forschungsziel der Studie lag darin, herauszufinden, ob es mit dem entwickelten Modell möglich ist, die erlebte Beanspruchung der Versuchspersonen während verschiedener Verkehrsszenarien zu replizieren. Die Versuchspersonen hatten hierbei die Aufgabe, den Verkehr, also Landungen und Starts, sowie das Rollen von Flugzeugen zur Startbahn und zur Parkposition an einem kleinen Flughafen zu koordinieren (näheres dazu in Kapitel 6.1.5). Da die Belastung für die Versuchspersonen sonst zu hoch geworden wäre, wurde außerdem entschieden, die als Nebenaufgabe anzusehende Verkehrsplanung (Annahme der Flugstreifen aus dem „Flugstreifendrucker“ sowie Einsortierung in die Stripbay) mittels der Flugstreifen

nicht umzusetzen. Sie hatten allerdings die zusätzliche Aufgabe, jede Minute den RSME-Fragebogen auszufüllen. Die Flugstreifen wurden daher bereits vor dem Szenario in der Stripbay vorsortiert, sodass das Notizen-machen auf den Flugstreifen und ihr Nutzen als Gedankenstütze dabei für die Versuchspersonen erhalten blieb.

Hinsichtlich der Versuchsdurchführung ist anzumerken, dass es bei zwei Versuchspersonen zu Fehlfunktionen des Simulators kam. Bei einer Versuchsperson stürzte die Simulation nach zwei Minuten ab, sodass das Szenario erneut begonnen werden musste. Bei einer zweiten Versuchsperson traten aufgrund der Fehlfunktion nur sechs der geplanten zehn Flugzeuge auf. Für die Einzelfallanalysen wurden diese Versuchspersonen aufgrund fehlender Beanspruchungsratings ohnehin aus der Untersuchung ausgeschlossen. Bei einer Überprüfung der Daten der beiden Versuchspersonen im Vergleich zur restlichen Stichprobe wurden jedoch keine Auffälligkeiten festgestellt, sodass für die Analyse der mittleren Beanspruchungshöhe beide Versuchspersonen in der Analyse belassen wurden. Ein weiterer Versuchsdurchlauf musste abgebrochen werden, da die Versuchsperson auch nach zwei kompletten Versuchsszenarien die Aufgabe nicht vollkommen verstanden hatte. Dieser Versuchsdurchlauf wurde vollkommen aus der Analyse ausgeschlossen und mit einer anderen Versuchsperson nacherhoben.

Als *unabhängige Variable* wurde das **Verkehrsaufkommen** (die Anzahl der Flugzeuge, die innerhalb des Szenarios starten, landen und rollen möchten) zweistufig variiert, wobei jede Versuchsperson beide Faktorstufen durchlief (Within-Subjects-Design). Hohes Verkehrsaufkommen wurde mit 20 Flugzeugen pro 15 Minuten (ca. 1,3 Bewegungen/Minute) bestimmt, niedriges Verkehrsaufkommen mit 10 Flugzeugen pro 15 Minuten (ca. 0,6 Bewegungen/Minute) [63]. Zusätzlich wurde für jede Faktorstufe eine Messwiederholung durchgeführt (insgesamt 3 Messwiederholungen für jede AV), so dass die Versuchspersonen vier 15-minütige Verkehrsszenarien zu bearbeiten hatten. Die Szenarien 1 und 3 hatten dabei ein hohes Verkehrsaufkommen, die Szenarien 2 und 4 ein geringes (vgl. Tabelle 3). Die Szenarien 1 und 2 unterschieden sich dabei lediglich in der Anzahl der abzufertigenden Flugzeuge voneinander, während die Reihenfolge des Auftauchens, sowie die Rufzeichen konstant gehalten wurden. Zur Erstellung der Szenarien 3 und 4 wurden aus den Szenarien 1 und 2 jeweils die Rufzeichen, sowie geringfügig die Verteilung der Flugzeuge über die Zeit verändert. Um

[63] Bei der Festlegung der Anzahl der Flüge pro Zeiteinheit wurde sich an realen Daten orientiert. Da das Ziel die Erzeugung einer hohen Beanspruchung war, wurde hohes Verkehrsaufkommen anhand der Werte des stark frequentierten Flughafens Frankfurt a. M. festgelegt, an dem 2009 etwa 83 Bewegungen (Starts und Landungen) pro Stunde die Regel waren (ca. 1,3 Bewegungen/Minute). Ab 2011 ist durch den Zubau einer weiteren Landebahn sogar 126 Bewegungen/Stunde (2,1 Bewegungen/Minute) erreicht worden (Huber, 2012). Geringes Verkehrsaufkommen wurde anhand der bei Huber (2012) gegebenen Werte von 36 Bewegungen pro Stunde festgelegt (0,6 Bewegungen/Minute). Diese Werte wurden zusätzlich in Vorversuchen überprüft.

Reihenfolgeeffekte auszuschließen und Lerneffekte besser kompensieren zu können wurden die Szenarien über alle Versuchspersonen permutiert.

Tabelle 3: Übersicht über die Szenarienbezeichnung und der mit ihnen einhergehenden Belastung

	Hohes Verkehrs-aufkommen	**Niedriges Verkehrs-aufkommen**
Gruppe 1	Szenario 1	Szenario 2
Gruppe 2	Szenario 3	Szenario 4

Dabei ergaben sich 4! = 24 mögliche Anordnungen der Szenarien hintereinander. Diese sind in Tabelle 4 dargestellt. So erhielt Versuchsperson 14 beispielsweise zuerst das Szenario 3 (hohes Verkehrsaufkommen), dann 4 (niedriges Verkehrsaufkommen), anschließend die Szenarien 1 (hohes Verkehrsaufkommen) und 2 (niedriges Verkehrsaufkommen).

Tabelle 4: Permutation der Szenarien für die 24 Versuchspersonen

Möglichkeit	**Reihenfolge der Verkehrsszenarien**	**Möglichkeit**	**Reihenfolge der Verkehrsszenarien**
1	1, 2, 3, 4	13	3, 4, 2, 1
2	1, 2, 4, 3	14	3, 4, 1, 2
3	1, 3, 2, 4	15	3, 2, 1, 4
4	1, 3, 4, 2	16	3, 2, 4, 1
5	1, 4, 2, 3	17	3, 1, 2, 4
6	1, 4, 3, 2	18	3, 1, 4, 2
7	2, 3, 4, 1	19	4, 3, 2, 1
8	2, 3, 1, 4	20	4, 3, 1, 2
9	2, 4, 3, 1	21	4, 2, 3, 1
10	2, 4, 1, 3	22	4, 2, 1, 3
11	2, 1, 3, 4	23	4, 1, 2, 3
12	2, 1, 4, 3	24	4, 1, 3, 2

Als *abhängige Variablen* wurden für die Erhebung der subjektiven mentalen Beanspruchung während der Szenarien die modifizierte Rating Scale Mentel Effort **RSMEmod** (Kosicki, 2011; im Folgenden nur mit RSME bezeichnet; siehe Anhang) verwendet. Hierbei waren die

Versuchspersonen aufgefordert, jede Minute eine Einschätzung ihrer aktuellen mentalen Beanspruchung auf der RSME-Skala abzugeben.

Zusätzlich wurde am Ende jedes Szenarios der **NASA-TLX** erfasst (siehe Anhang), in dem die Versuchspersonen ihre Beanspruchung über das gesamte Szenario einschätzen sollten. Hiermit sollte nochmals kontrolliert werden, ob die Manipulation im Experiment geglückt ist, sprich, ob die Versuchspersonen die Szenarien mit hohem Verkehrsaufkommen als stärker beanspruchend empfanden als diejenigen Szenarien mit geringem Verkehrsaufkommen. Weiterhin bestand hierdurch die Möglichkeit eines Vergleiches zwischen den Bewertungen der mittleren Beanspruchung des NASA-TLX und der RSME-Skala, um die höhere Diagnostizität multidimensionaler Skalen auszunutzen.

Als Leistungsmaß wurde außerdem die **Fehlerzahl** herangezogen, zu deren Ermittlung Funkprotokolle aufgenommen wurden, aus denen die Fehler herausgelesen werden konnten. Als Fehler wurden dabei folgende Aktionen und Ereignisse eingestuft:

- Verzögerungen, die nicht durch die Verkehrssituation induziert waren: z.B. wenn nach einer Anfrage zu lang gewartet wird, um die Freigabe zu erteilen, obwohl die Verkehrssituation es zulässt, diese früher zu erteilen.
- Zu geringe Staffelungen auf den Rollwegen: d.h., wenn durch eine zu enge Erteilung von Freigaben, die Flugzeuge zu dicht hintereinander herrollen.
- Freigabeverwechslungen/falsche Freigaben: wenn z.B. statt „rollen zum Terminal“ „rollen zur Piste“ gegeben wird; es folgt darauf keine Aktion des Flugzeuges, sondern eine Wiederholung der Anfrage.
- Rufzeichen- und/oder Flugzeugverwechslung: Der Proband gibt eine Freigabe für Flugzeug Y, meint aber das Flugzeug X.
- Keine Freigabeerteilung: Auf Anfragen von Flugzeugen wird schlichtweg nicht reagiert, obwohl es die Verkehrssituation zulässt. Landende Flugzeuge müssen durchstarten, Flugzeuge am Gate können dieses nicht verlassen und es kommt zu starken Verzögerungen.
- Zusammenstoß in der Luft: Benötigt einen vorherigen Fehler des Probanden, dass die Landefreigabe nicht erteilt wird; dadurch startet das landende Flugzeug durch und es kann passieren, dass es mit einem gerade startenden Flugzeug in der Luft kollidiert.
- Stau an der Start-/Landebahn: Vor dem Haltepunkt ist gerade so viel Platz, dass zwei Flugzeuge hintereinander an der Startbahn auf eine Startfreigabe warten können. Warten mehr als zwei Flugzeuge, so stauen sich die Wartenden auf den Hauptrollweg zurück, welcher für das Rollen zum Terminal anderer Flugzeuge benötigt wird. Hierdurch kommt es zu starken Verzögerungen.

- Vorziehen eines Starts, obwohl ein Flugzeug eine Landeanfrage gestellt hat: Laut Instruktion sind Landungen Starts immer vorzuziehen. Starts können nur erfolgen, wenn die Verkehrssituation es zulässt d.h. wenn genügend Abstand zwischen dem startenden und dem landenden Flugzeug gewährleistet ist und sie sich nicht gleichzeitig auf der Start-/ Landebahn befinden.
- Staffelungsunterschreitung auf der Start-/Landebahn: Es befinden sich gleichzeitig zwei Flugzeuge auf der Start-/Landebahn, stoßen aber nicht zusammen.
- Runway Incursion: Zwei Flugzeuge stoßen auf den Rollwegen oder der Start-/Landebahn zusammen; z.B. die Freigabe wird so gegeben, dass ein Flugzeug von der Parkposition auf den Rollweg aufrollt, obwohl sich dort gerade ein anderes Flugzeug befindet.
- Proaktive Freigabeerteilung: Der Proband wartet die Anfrage von Flugzeugen nicht ab, sondern erteilt eine Freigabe an ein Flugzeug bereits bevor dieses Kontakt zu ihm aufgenommen hat.

Das Auftreten von Fehlern hatte dabei für die Versuchspersonen keine gravierenden Auswirkungen. Es kam lediglich zu leichten Verzögerungen bzw. dazu, dass sich z.B. bei Zusammenstößen kurzzeitig zwei Flugzeugsymbole auf dem Radardisplay überlagerten. Die Simulation lief ansonsten ganz normal weiter.

Zusätzlich zu den subjektiven Ratingskalen wurde während der gesamten Szenarien als physiologisches Beanspruchungsmaß das **EKG** der Versuchspersonen gemessen, woraus die Herzrate, gemessen als Länge der Abstände zwischen zwei Herzschlägen (Inter Beat Intervalle in ms; ms IBI), extrahiert wurde, um einen weiteren Beanspruchungsindikator zu erhalten. Eine Fragestellung ist hierbei, ob sich die subjektiven Ratings von den objektiven EKG Daten unterscheiden, d.h. die EKG-Daten eventuell eine andere Verteilung der Beanspruchung erkennen lassen. Außerdem sollte zur Testung der Modellgüte eine Korrelation der EKG-Daten mit den Modelldaten erfolgen. Nicht zuletzt ist es daher auch interessant zu vergleichen, wie gut die Modelldaten mit den subjektiven Fragebogendaten und den objektiven EKG Daten korrelieren und ob es Unterschiede gibt (vgl. Kapitel 6.1.4).

Als *Kontrollvariable* wurde außerdem ein **d2-R Aufmerksamkeitstest** durchgeführt (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010). Der Test dient der Messung der Konzentration bei Aufgaben, die Aufmerksamkeit verlangen (konzentrierte Aufmerksamkeit). Er erfasst die Konzentrationsfähigkeit der Testperson sowie die Schnelligkeit und Genauigkeit bei der Unterscheidung ähnlicher visueller Reize (Detail-Diskrimination). Dies diente der Erhebung einer Kontrollvariablen für die Fähigkeit der Versuchspersonen, schwierige Aufgaben schnell und gründlich zu bearbeiten. Da anzunehmen ist, dass das Abschneiden im d2-R Test und die damit verbundene Schnelligkeit und Genauigkeit der Aufgabenbearbeitung einen Einfluss auf die

Fehlerzahl und die Bewertung der mentalen Beanspruchung hat, ist eine Testung angedacht, in wieweit die Versuchspersonen anhand ihres d2-R Ergebnisses in Gruppen eingeteilt werden können, die sich signifikant hinsichtlich ihrer Fehlerzahl bzw. der abgegebenen Beanspruchungseinschätzung unterscheiden (vgl. Kapitel 6.1.4).

6.1.4 HYPOTHESEN

Die zur Untersuchung stehenden Hypothesen betreffen jeweils das verwendete Messinstrument für mentale Beanspruchung und sind ihrerseits nochmals in Unterhypothesen unterteilt. Es wurden zunächst zwei Hypothesen formuliert, von denen das weitere Vorgehen zur Auswertung der Daten abhängt. Einerseits galt es feststellen, ob ein Einfluss des Abschneidens im d2-R-Test auf die Fehlerzahl und die Beanspruchungsratings vorliegt. Aufgrund der unklaren Befundlage zur Herzrate sollte im Weiteren untersucht werden, ob im vorliegenden Experiment mit der Herzrate eine Unterscheidung zwischen den Belastungsbedingungen möglich ist oder ob der Effekt bestätigt wurde, dass lediglich die Experimentalbedingungen von der Baseline-Bedingung zu unterscheiden ist (Nickel et al., 2002). Beide Hypothesen haben Auswirkungen auf die dann folgende Vorgehensweise bei der Auswertung der übrigen Hypothesen.

So beschäftigt sich die **erste Hypothese** mit dem Zusammenhang zwischen dem Abschneiden im d2-R-Test und der Höhe der mittleren Fehlerzahl bzw. der mittleren Beanspruchungsratings. Ziel ist es dabei festzustellen, ob sich die Versuchspersonen anhand ihres d2-R-Wertes in Gruppen einteilen lassen, die sich signifikant in der Anzahl der gemachten Fehler, sowie der angegebenen Beanspruchungshöhe unterscheiden. Allgemein ist davon auszugehen, dass die Gründlichkeit und Effizienz bzw. der Effizienz-Gründlichkeits-Tradeoff, mit dem Versuchspersonen kognitive Aufgaben bearbeiten, einen Einfluss darauf hat, wie hoch die Beanspruchung bei der Aufgabenbearbeitung eingeschätzt wird. Tendiert die Versuchsperson beispielsweise dazu, sehr gründlich zu überlegen, so schlägt sich dies in einer langsamen Bearbeitungszeit nieder (im d2-R durch die Anzahl bearbeiteter Zielobjekte (BZO) erfasst). Übertragen auf die Aufgaben der Flugverkehrskontrolle bedeutet dies, dass Freigaben ggf. zu spät erteilt werden und es damit zu Konflikten bzw. Verzögerungen kommen kann. Auf der anderen Seite bedeutet ein sehr schnelles Abarbeiten, dass Zielobjekte übersehen werden (im d2-R wiedergegeben durch die Anzahl an Auslassungs- und Verwechslungsfehlern (AF und VF)). Es ist daher davon auszugehen, dass mit steigendem d2-R-Wert der Versuchsperson (besserem Abschneiden im Test), die Fehlerzahl abnimmt. Gleichzeitig ist anzunehmen, dass der Tradeoff zwischen Effizienz und Gründlichkeit weniger Aufwand benötigt, wodurch auch die Beanspruchung mit steigendem d2-R-Wert als geringer eingeschätzt werden sollte **(Hypothese 1)**.

Hypothese (1): Versuchspersonen mit besserem Abschneiden im d2-R-Test, machen auch in der Fluglotsenaufgabe im Mittel weniger Fehler und schätzen die mentale Beanspruchung (RSME und NASA-TLX) im Mittel geringer ein.

a. Mit steigendem d2-R-Wert werden weniger Fehler gemacht.

b. Mit steigendem d2-R-Wert wird die Beanspruchung als geringer eingeschätzt.

Ergibt sich das in Hypothese 1 erwartete Bild, so ist für die weiteren Auswertungsschritte zur Fehlerzahl und den Beanspruchungsratings jeweils der Einfluss des Abschneidens im d2-R-Test als Kontrollvariable herauszurechnen. Gilt die Nullhypothese (es gibt keinen Einfluss des d2-R-Tests), so kann davon ausgegangen werden, dass kein Einfluss vorliegt und der d2-R-Wert für die weitere Auswertung nicht berücksichtigt werden muss.

Die **zweite Hypothese** beschäftigt sich mit den Befunden zur Sensitivität der Herzrate. Aufgrund der unklaren Befundlage zur Herzrate gilt es daher zu testen, ob sich die beiden Experimentalbedingungen (hohes vs. niedriges Verkehrsaufkommen) signifikant voneinander unterscheiden oder ob nur die Baseline-Bedingung von den beiden Experimentalbedingungen zu unterscheiden ist (Nickel et al., 2002). Da die Herzrate im vorliegenden Fall nicht in Beats Per Minute, sondern die Länge der Inter-Beat-Intervalle in Millisekunden gemessen wurde, ergibt sich folgender Zusammenhang: Es ist davon auszugehen, dass eine Erhöhung der Belastung und damit der mentalen Beanspruchung mit einer Erhöhung der Herzrate (in bpm) einhergeht. Eine erhöhte Herzrate ergibt eine erhöhte Anzahl an Beats Per Minute, wodurch sich die Abstände zwischen den einzelnen Herzschlägen verringern, was zu einer geringeren Dauer der Inter-Beat-Intervalle (in ms) führt. Hierbei gilt, dass die Dauer der Inter-Beat-Abstände bei hohem Verkehrsaufkommen am geringsten, in der Baseline-Bedingung am größten ist **(Hypothese 2)**.

Hypothese (2): Die Herzrate der beiden Experimentalbedingungen (hohes vs. niedriges Verkehrsaufkommen), gemessen in Millisekunden Dauer der Inter-Beat-Intervalle, ist signifikant geringer als die der Baseline-Bedingung. Die mittleren IBI der Szenarien mit hohem Verkehrsaufkommen (1 und 3) sind geringer, als die mittleren IBI der Szenarien mit geringem Verkehrsaufkommen (2 und 4).

a. Die mittleren IBI der Szenarien mit hohem Verkehrsaufkommen (1 und 3) sind geringer als die der Baseline Bedingung.

b. Die mittleren IBI der Szenarien mit geringem Verkehrsaufkommen (2 und 4) sind geringer als die der Baseline Bedingung.

c. Die mittleren IBI der Szenarien mit hohem Verkehrsaufkommen (1 und 3) sind geringer als die der Szenarien mit geringem Verkehrsaufkommen (2 und 4).

Zeigt sich, dass die Unterscheidung zwischen den beiden Experimentalbedingungen möglich ist, so kann im Weiteren davon ausgegangen werden, dass sich die Experimentalbedingungen hinreichend stark voneinander unterscheiden, so dass die Sensitivität der Herzrate ausreicht, um auch Unterschiede zwischen den Szenarien festzustellen. Es ist daher mit einer Verringerung der Dauer der Inter-Beat-Intervalle bei steigender mentaler Beanspruchung zu rechnen **(Hypothese 3)**. Hierbei sollte sich zeigen, dass die Inter-Beat-Abstände bei hoher Belastung (hohes Verkehrsaufkommen) geringer sind, als bei geringer Belastung (geringes Verkehrsaufkommen). Somit sollten die Versuchspersonen für die Szenarien mit hohem Verkehrsaufkommen (Szenarien 1 und 3) geringere mittlere Inter-Beat-Abstände aufweisen, als in Szenarien mit niedrigem Verkehrsaufkommen (Szenarien 2 und 4). Die Inter-Beat-Abstände der beiden Szenarien mit hohem Verkehrsaufkommen (1 und 3) und mit niedrigem Verkehrsaufkommen (2 und 4) untereinander sollten sich allerdings nicht signifikant unterscheiden. Zeigt sich in Hypothese 2, dass keine Unterscheidung der Experimentalbedingungen möglich ist, so ist die Herzrate aus der weiteren Auswertung auszuschließen, da kein Informationsgewinn zu erwarten ist.

Hypothese (3): Die Herzrate, gemessen in Millisekunden Dauer der Inter-Beat-Intervalle, ist bei hohem Verkehrsaufkommen signifikant geringer als bei niedrigem Verkehrsaufkommen (3a – d). Dagegen unterscheidet sich die Herzrate von Szenarien mit gleich hohem Verkehrsaufkommen nicht (3e – f).

a. Die mittleren IBI sind in Szenario 1 geringer als in Szenario 2.

b. Die mittleren IBI sind in Szenario 1 geringer als in Szenario 4.

c. Die mittleren IBI sind in Szenario 3 geringer als in Szenario 2.

d. Die mittleren IBI sind in Szenario 3 geringer als in Szenario 4.

e. Die mittleren IBI der Szenarien 1 und 3 unterscheiden sich nicht.

f. Die mittleren IBI der Szenarien 2 und 4 unterscheiden sich nicht.

Hierbei ist darauf hinzuweisen, dass bei den Hypothesen 3e und 3f die Forschungshypothese als Nullhypothese formuliert wurde, die Hypothesen 3a – d als Alternativhypothesen. Dies ist beim Setzen des α-Niveaus zu beachten. So wurde das α-Niveau der Alternativhypothesen traditionell auf $\alpha = 0{,}05$ gesetzt. Für den Fall, dass Nullhypothesen als Forschungshypothesen verwendet werden, plädieren Bortz und Döring (2006) dafür das α-Niveau mindestens auf $\alpha = 0{,}10$ festzulegen, um die Teststärke zu erhöhen und somit das Risiko eines β-Fehlers zu senken.

Bortz (2005) plädiert sogar für ein Niveau von $\alpha = 0{,}20$, was auch hier verwendet werden soll, wobei zweiseitig getestet wird[64].

Die **vierte Hypothese** beschäftigt sich mit der subjektiven Beurteilung der mentalen Beanspruchung, die durch die RSME-Skala erhoben wurde. Hierbei wurde davon ausgegangen, dass es zu höheren Beanspruchungs-Ratings kommt, wenn die Belastung durch die Erhöhung der Anzahl der Flugzeuge zunimmt. Durch die zusätzlich zu bearbeitenden Flugzeuge müssen mehr Informationen im Arbeitsgedächtnis behalten werden, um adäquate Entscheidungen zu treffen, was zu einer höheren Belastung der kognitiven Ressourcen und damit zu höheren Beurteilungen in den RSME-Skalen führt **(Hypothese 4)**. Hierbei sollte sich zeigen, dass die Beanspruchung bei hoher Belastung (hohes Verkehrsaufkommen) höher eingeschätzt wird als bei geringer (geringes Verkehrsaufkommen). Somit sollten die Versuchspersonen für die Szenarien mit hohem Verkehrsaufkommen (Szenarien 1 und 3) die mentale Beanspruchung als signifikant höher einschätzen, als in Szenarien mit niedrigem Verkehrsaufkommen (Szenarien 2 und 4). Die Bewertungen der mentalen Beanspruchung der beiden Szenarien mit hohem Verkehrsaufkommen (1 und 3) und mit niedrigem Verkehrsaufkommen (2 und 4) untereinander sollten sich allerdings nicht signifikant unterscheiden. Daraus ergeben sich die folgenden sechs Unterhypothesen:

Hypothese (4): Die subjektiven Bewertungen der mentalen Beanspruchung, erhoben mit der RSME-Skala, sind bei hohem Verkehrsaufkommen signifikant höher als bei niedrigem (4a – d). Die Bewertungen von Szenarien mit gleich hohem Verkehrsaufkommen unterscheiden sich nicht voneinander (4e – f).

a. Die subjektive Bewertung ist in Szenario 1 höher als in Szenario 2.

b. Die subjektive Bewertung ist in Szenario 1 höher als in Szenario 4.

c. Die subjektive Bewertung ist in Szenario 3 höher als in Szenario 2.

d. Die subjektive Bewertung ist in Szenario 3 höher als in Szenario 4.

e. Die subjektiven Bewertungen der Szenarien 1 und 3 unterscheiden sich nicht.

f. Die subjektiven Bewertungen der Szenarien 2 und 4 unterscheiden sich nicht.

[64] Ein hohes α-Niveau sorgt dafür, dass nicht signifikante Ergebnisse unwahrscheinlicher werden und damit die Annahme der H_0 erschwert wird. Es ist weiterhin darauf hinzuweisen, dass auch bei nicht signifikanten Ergebnissen und einer damit verbundenen Annahme der Nullhypothese, diese jedoch nicht als „bestätigt" anzusehen ist. Auch bei hoher Teststärke bedeutet ein nicht signifikantes Ergebnis, dass die H_1 mit niedriger β-Fehlerwahrscheinlichkeit fälschlicherweise abgelehnt wird (Bortz & Döring, 2005). Da die Nullhypothese besagt, dass es keinen Effekt gibt, bedeutet ein nicht signifikantes Ergebnis aber nicht, dass kein Effekt vorliegt. Es bedeutet lediglich, dass der Effekt so klein ist, dass er durch Zufall zustande gekommen sein könnte (Field, 2009).

Auch hier gilt für die beiden als Nullhypothese formulierten Forschungshypothesen 4e und 4f das erhöhte α-Niveau von α = 0,20.

Für die Erhebung der mentalen Beanspruchung mit Hilfe des NASA-TLX wurde in **Hypothese fünf** analog verfahren wie oben und dieselben Hypothesen geprüft.

Hypothese (5): Die subjektiven Bewertungen der mentalen Beanspruchung, erhoben mit dem NASA-TLX (Sub-Skala geistige Anstrengung), sind bei hohem Verkehrsaufkommen signifikant höher als bei niedrigem (5a – d). Die Bewertungen von Szenarien mit gleich hohem Verkehrsaufkommen unterscheiden sich nicht voneinander (5e – f).

a. Die subjektive Bewertung ist in Szenario 1 höher als in Szenario 2.

b. Die subjektive Bewertung ist in Szenario 1 höher als in Szenario 4.

c. Die subjektive Bewertung ist in Szenario 3 höher als in Szenario 2.

d. Die subjektive Bewertung ist in Szenario 3 höher als in Szenario 4.

e. Die subjektiven Bewertungen der Szenarien 1 und 3 unterscheiden sich nicht.

f. Die subjektiven Bewertungen der Szenarien 2 und 4 unterscheiden sich nicht.

Auch hier gilt für die beiden als Nullhypothese formulierten Forschungshypothesen 5e und 5f das erhöhte α-Niveau von α = 0,20.

Als weitere abhängige Variable wurde als Leistungsmaß die Fehlerzahl erhoben, welche in **Hypothese sechs** untersucht wird. Es wird hierbei davon ausgegangen, dass bei höherer Belastung (Anzahl der abzufertigen Flugzeuge) auch die Belastung des Arbeitsgedächtnisses zunimmt. Dies führt dazu, dass einerseits mehr Vergessensfehler auftreten (es werden Flugzeuge nicht oder zu spät abgefertigt; schlichtweg vergessen). Andererseits wird angenommen, dass es vermehrt zu schlechten Entscheidungen kommt, die zu kritischen Situationen (Zusammenstößen, Durchstarten von Flugzeugen) führen. Hierbei ist erneut davon auszugehen, dass sich auch die Fehlerzahlen der Szenarien mit hohem Verkehrsaufkommen (Szenarien 1 und 3) von den Fehlerzahlen der Szenarien mit niedrigem Verkehrsaufkommen (Szenarien 2 und 4) signifikant voneinander unterscheiden. Hingegen sollten sich die Fehlerzahlen der beiden Szenarien mit hohem Verkehrsaufkommen (1 und 3) und der beiden Szenarien mit niedrigem Verkehrsaufkommen (2 und 4) untereinander nicht signifikant unterscheiden.

Hypothese (6): Die Fehlerzahlen sind bei hohem Verkehrsaufkommen signifikant höher als bei niedrigem. Die Fehlerzahlen von Szenarien mit gleich hohem Verkehrsaufkommen unterscheiden sich nicht signifikant voneinander.

a. Die mittlere Fehlerzahl ist in Szenario 1 höher als in Szenario 2.

b. Die mittlere Fehlerzahl ist in Szenario 1 höher als in Szenario 4.

c. Die mittlere Fehlerzahl ist in Szenario 3 höher als in Szenario 2.

d. Die mittlere Fehlerzahl ist in Szenario 3 höher als in Szenario 4.

e. Die mittlere Fehlerzahl ist der Szenarien 1 und 3 unterscheiden sich nicht.

f. Die mittlere Fehlerzahl ist der Szenarien 2 und 4 unterscheiden sich nicht.

Auch hier gilt für die beiden als Nullhypothese formulierten Forschungshypothesen 6e und 6f das erhöhte α-Niveau von α = 0,20. Alle zur Überprüfung der Experimentellen Manipulation aufgestellten Hypothesen sind nochmals in Tabelle 5 zusammengefasst.

Tabelle 5: Übersicht der Hypothesen zur Überprüfung der experimentellen Manipulation und Erhebung von Validierungswerten

	Nr.	Hypothese
Versuchspersonen (experimentelle Manipulation)	Einfluss d2-R	
	1	Ein höherer d2-R-Wert führt zu weniger Fehlern und geringeren Beanspruchungswerten. a. Mit steigendem d2-R-Wert werden weniger Fehler gemacht. b. Mit steigendem d2-R-Wert wird die Beanspruchung als geringer eingeschätzt
	Physiologisches Maß	
	2	Herzrate (ms IBI) sinkt mit steigender Belastung. Baseline > low load > high load
	3	Die Herzrate (ms IBI) ist bei hohem Verkehrsaufkommen geringer als bei niedrigem Verkehrsaufkommen. SZN 1, 3 < SZN 2, 4; SZN 1 = SZN 3; SZN 2 = SZN 4
	Subjektive Maße	
	4	Die RSME-Ratings sind bei hohem Verkehrsaufkommen höher als bei niedrigem. SZN 1, 3 > SZN 2, 4; SZN 1 = SZN 3; SZN 2=SZN 4
	5	Die NASA-TLX Ratings (Sub-Skala „geistige Anstrengung“) sind bei hohem Verkehrsaufkommen höher als bei niedrigem.

		SZN 1, 3 > SZN 2, 4; SZN 1 = SZN 3; SZN 2=SZN 4
	Leistungsmaß	
	6	Die Fehlerzahl ist bei hohem Verkehrsaufkommen höher als bei niedrigem. SZN 1, 3 > SZN 2, 4; SZN 1 = SZN 3; SZN 2 = SZN 4

Neben der Abbildung der Befunde zum Einfluss des Verkehrsaufkommens auf die mentale Beanspruchung ist das Hauptziel der Studie die **Überprüfung der Modellgüte**. Diese erfolgt durch den Vergleich zwischen realen (hier experimentell ermittelten) Beanspruchungsdaten mit den Beanspruchungsdaten des Modells. Die Beurteilung der Güte eines Modells erfolgt gemeinhin durch die Berechnung sogenannter Goodness-of-Fit-Maße. Schunn und Wallach (2005) beschreiben eine Reihe solcher Maße. Die Berechnung dieser Goodness-of-Fit-Maße erfolgt meist durch einen Vergleich des Verlaufes der aggregierten Daten (Experimental- und Modelldaten) über die Zeit. Da eine Standardisierung der Szenarien auf Grund der Beschaffenheit der Simulationssoftware nicht möglich war, ist eine solche Berechnung eines mittleren zeitlichen Verlaufes aus den Rohwerten des RSME sowie der Modellwerte nicht angebracht, sodass Einzelfallanalysen durchgeführt werden. Daher soll die Überprüfung und Berechnung dieser Maße im vorliegenden Fall für jede Versuchsperson individuell durchgeführt werden. Dies bringt zwar den Nachteil mit sich, dass eine Menge Rauschen auf den Daten liegt, da individuelle Unterschiede und Ausreißer stärker ins Gewicht fallen, hat aber den Vorteil, dass der Fit des Modells anhand jedes Einzelfalles ermittelt werden kann. Einzelfallanalysen bieten darüber hinaus sehr viel mehr Aufschluss über bestimmte Muster der Daten und individuelle Verhaltensweisen und/oder Eigenschaften der Versuchspersonen (Sarris & Reiß, 2005; Thorngate & Edmonds, 2012). Zusätzlich kann auf Grund der Einzelfallanalysen bei einem guten Fit der Modelldaten auf die Experimentaldaten von einem starken Modell gesprochen werden, da trotz des Rauschens ein guter Fit nachgewiesen werden kann.

Schunn und Wallach (2005) beschreiben drei „Säulen“ der Goodness-Of-Fit-Prüfung: Einen qualitativen visuellen Vergleich der Verläufe, sowie die quantitative Berechnung von Maßen des relativen Trends und der Abweichung vom exakten Punkt. Vor dem Hintergrund dessen, was das Modell replizieren soll und kann, soll zunächst eine **visuelle Überprüfung des Fits** angewendet werden (Schunn & Wallach, 2005; Thorngate & Edmonds, 2012). Dieser kann für den vorliegenden Versuch zum einen durch einen Vergleich der mittleren Beanspruchungshöhe der subjektiven Daten mit den Modelldaten innerhalb der einzelnen Szenarien erfolgen; zum anderen durch visuelle Überprüfung der Ähnlichkeit von Verläufen von Modell- und experimentellen Daten. Letzterer soll für jede Versuchsperson individuell durchgeführt werden

(Einzelfallbetrachtungen). Hierbei wird einerseits durch Streudiagramme ein visueller Eindruck über den Zusammenhang der Modell- und Experimentaldaten gewonnen. Auf der anderen Seite kann anhand der zeitlichen Verläufe der abhängigen Variablen RSME-Bewertung, ms IBI, sowie der Modelldaten (Anzahl der Chunks im Arbeitsgedächtnis) festgestellt werden, wie gut gewisse Trends, sowie Gipfel und Täler von Modell- und Versuchspersonendaten übereinstimmen. Dabei ist es vor allem von Interesse, herauszufinden, ob das Modell Beanspruchungs-Gipfel replizieren kann, da ein wesentliches Ziel des Modells die Aufdeckung solcher Zeiträume hoher Beanspruchung darstellt.

Ferner kann statistisch überprüft werden, ob sich für die mittlere Beanspruchungshöhe in jedem Szenario ein ähnliches Bild ergibt, wie bereits bei den Hypothesen zu den subjektiven Bewertungen und dem Leistungsmaß dargelegt wurde: Eine höhere Belastung sollte mit höheren Beanspruchungswerten des Modells einhergehen, geringere Belastung im Umkehrschluss mit geringeren Beanspruchungswerten. Es ergibt sich **Hypothese 7** äquivalent zu den Hypothesen 4, 5 und 6:

Hypothese (7): Die mittleren Beanspruchungswerte des Modells (gemessen über die Anzahl der Chunks im Arbeitsgedächtnis des Modells) sind bei hohem Verkehrsaufkommen signifikant höher als bei niedrigem. Die mittleren Beanspruchungswerte des Modells von Szenarien mit gleich hohem Verkehrsaufkommen unterscheiden sich nicht voneinander.

a. Die mittlere Beanspruchungshöhe des Modells ist in Szenario 1 höher als in Szenario 2.

b. Die mittlere Beanspruchungshöhe des Modells ist in Szenario 1 höher als in Szenario 4.

c. Die mittlere Beanspruchungshöhe des Modells ist in Szenario 3 höher als in Szenario 2.

d. Die mittlere Beanspruchungshöhe des Modells ist in Szenario 3 höher als in Szenario 4.

e. Die mittleren Beanspruchungshöhen des Modells der Szenarien 1 und 3 unterscheiden sich nicht.

f. Die mittleren Beanspruchungshöhen des Modells der Szenarien 2 und 4 unterscheiden sich nicht.

Wiederum gilt hier für die beiden als Nullhypothese formulierten Forschungshypothesen 7e und 7f das erhöhte α-Niveau von $\alpha = 0{,}20$.

Ferner sollen verschiedene quantitative Goodness-of-Fit-Maße berechnet werden, um die Güte des Modells quantitativ beschreiben zu können. Schunn und Wallach (2005) beschreiben zwei Arten von Goodness-of-Fit-Maßen: Einerseits Maße, die die Abweichung von einem exakten Punkt wiedergeben und andererseits Maße, die den relativen Trend wiedergeben. Jedes Maß deckt hierbei immer nur eine der beiden Informationen ab (Schunn & Wallach, 2005). Als **Maße**

des relativen Trends schlagen sie die Anwendung von Pearson's r und r^2 vor, falls Verhältnisskalenniveau vorliegt. Da dieses bei den erhobenen Fragebogendaten nicht angenommen werden kann, weil die Skala u.a. nicht äquidistant ist (Annett, 2002; Bortz & Lienert, 2008; Fleid, 2009; Schunn & Wallach, 2005), schlagen Schunn und Wallach (2005) die Verwendung der Rangkorrelationskoeffizienten Spearman's ρ und Kendall's τ vor. Spearman's ρ gleicht dabei in der mathematischen Berechnung Pearson's r und ist damit mit r vergleichbar. Die Berechnung beruht allerdings nicht auf den tatsächlichen Messwerten, sondern auf deren Rängen. Spearman's ρ ist außerdem sinnvoll, weil die abhängige Variable des Modells (in diesem Fall Anzahl der Chunks im Arbeitsgedächtnis) nur sehr lose mit der abhängigen Variable der Realdaten (RSME-Bewertung) verwandt ist (Schunn & Wallach, 2005). Ferner wird ρ häufiger angewendet und gewährleistet so Vergleichbarkeit zwischen verschiedenen Befunden.

Als zweiter Kennwert kann Kendall's τ berechnet werden, welcher auf einem anderen Berechnungsverfahren beruht als Pearson's r und Spearman's ρ. Auch wenn Schunn und Wallach (2005) keinen Grund dafür sehen, dass Kendall's τ Spearman's ρ vorgezogen werden sollte, so soll dieser hier doch berechnet werden. Dies ist in der Tatsache begründet, dass Kendall's τ als weitaus mächtiger angesehen werden kann als andere Korrelationskoeffizienten, da alle Wertepaare untereinander verglichen werden und nicht nur jeweils die zwei Wertepaare eines Paares. Hierdurch ist τ meist etwas konservativer und somit die Gefahr Ergebnisse überzuinterpretieren damit etwas eingegrenzt. Im Gegensatz zu ρ wird anstatt der quadrierten Rangdifferenzen die Fehlordnung der Paare untereinander ausgewertet, wodurch Kendall's τ unempfindlicher gegenüber Ausreißern ist (Newton, 2002).

Eine weitere Möglichkeit, den Fit von Modell- und Realdaten speziell von ordinal skalierten Variablen zu berechnen, entwickelte Thorngate (1987; vgl. auch Thorngate & Edmonds, 2012, 2013). In Anlehnung seine Methode der *Ordinal Pattern Analysis* (OPA) kann eine Auswertung des relativen Trends der Daten erfolgen. So wird eine Kennzahl (der sogenannte Index of Observed Fit; IOF) erhalten, die aussagt, wie oft der Trend von einem Datenpunkt zum anderen Datenpunkt in den Modelldaten den Trend vom selben Datenpunkt zum nächsten in den Realdaten abbildet. Problematisch ist hierbei allerdings anzumerken, dass die Änderung der RSME-Daten innerhalb kleiner Bereiche liegen kann (wenige Zentimeter), was vom Modell nicht repliziert wird. Das bedeutet, dass die Änderung der Bewertungshöhe einer Versuchsperson von einem Zeitpunkt zum nächsten sehr klein sein kann, während das Modell solche feinen Unterschiede nicht abbilden kann (konkret bedeutet das, der Modellwert von Zeitpunkt 1 zu Zeitpunkt 2 würde gleich bleiben). Es kann zudem nicht ausgeschlossen werden, dass diese Differenz in der subjektiven Bewertung zufällig zustande kommt. Zusätzlich ist bei einem Scope von 1, also wenn für jeden Messwert auch ein Modellwert vorliegt, der Index of Observed Fit

äquivalent zu Kendall's τ (vgl. Thorngate & Edmonds, 2012), welcher ohnehin berechnet wird, so dass kein Gewinn an Informationen zustande kommt.

Es soll daher zunächst ein **Vergleich der Trends der mittleren Beanspruchungshöhe** von Modelldaten und subjektiven Daten bzw. Leistungsmaß über die vier Szenarien durchgeführt werden. Verglichen über die vier Szenarien sollte das Modell denselben Trend widerspiegeln, wie die subjektiven Maße und das Leistungsmaß (starke Korrelation (> 0,6); **Hypothese 8**).

Hypothese (8): Es besteht eine starke Korrelation (ρ; τ > 0,6) zwischen der mittleren Beanspruchungshöhe der subjektiven (RSME und NASA-TLX) und der Modelldaten über die vier Szenarien.

Weiterhin können jeweils die **zeitlichen Verläufe**, also die Einschätzungen der mentalen Beanspruchung mittels RSME eines Szenarios über die 15 Messzeitpunkte mit den dazugehörigen Modelldaten verglichen werden. Bei vier Szenarien pro Versuchsperson ergeben sich somit vier Korrelationskoeffizienten mit den Modellwerten, bei 24 Versuchspersonen also 96 Korrelationskoeffizienten. Der mittlere Korrelationskoeffizient eines Szenarios wird anschließend aus den erhaltenen individuellen Korrelationskoeffizienten für jede Versuchsperson mit Hilfe einer Fischer Z-Transformation ermittelt (Bortz & Schuster, 2010). Es wird hierbei ein positiver Zusammenhang zwischen den Modelldaten und den Bewertungen der mentalen Beanspruchung erwartet **(Hypothese 9)**. Das bedeutet, wenn die Beanspruchungsbewertung der Versuchspersonen steigt, gleichermaßen auch die Modellwerte steigen. Sinken die Beanspruchungsbewertungen der Versuchspersonen, so sinken auch die Modellwerte.

Hypothese (9): Es besteht ein positiver Zusammenhang zwischen den zeitlichen Verläufen der RSME- und Modelldaten.

Für den Fall, dass die Hypothesen 2 und 3 bestätigt werden können, können darüber hinaus in gleicher Weise **korrelative Zusammenhänge zwischen Modell und der Herzrate** (in ms IBI) untersucht werden. Hier ist allerdings der Korrelationskoeffizient Pearson's r anwendbar, da sowohl die Modelldaten, als auch die Herzdaten intervallskaliert sind. Es ist zu erwarten, dass ein negativer Zusammenhang zwischen den Modelldaten und der Herzrate, gemessen in ms IBI, besteht **(Hypothese 10)**. Das bedeutet, dass bei steigenden Inter-Beat-Intervallen sinkende Modellwerte zu erwarten sind. Gleichermaßen sind bei sinkenden Inter-Beat-Intervallen steigende Modellwerte zu erwarten.

Hypothese (10): Es besteht ein negativer Zusammenhang zwischen den zeitlichen Verläufen der Herzrate (in ms IBI) und den Modelldaten.

Als zweite Art der Godness-of-Fit-Maße beschreiben Schunn und Wallach (2005) die **Maße der Abweichung von einem exakten Punkt**. Hierbei soll sichergestellt werden, dass nicht nur der relative Trend eines Datensatzes, sondern auch die Größe der individuellen Messwerte hinreichend gut durch das Modell abgebildet wird. Selbst wenn der Trend eines Datensets gut abgebildet wird, so können die absoluten Werte um einige Größenordnungen voneinander abweichen, was einen schlechten Fit bedeuten würde. Dies bedeutet, dass entweder der Anstieg oder die Konstante einer Regressionsfunktion, die die Modelldaten auf die experimentellen Daten anpasst, abweichen (Schunn & Wallach, 2005). Als Maße werden hier meist die *mittlere quadratische Abweichung (MSD; mean squared deviation)* oder die *Wurzel der mittleren quadratischen Abweichung (RMSD; root mean squared deviation)* verwendet. Hierbei ist allerdings darauf zu achten, ob die Berechnung solcher Kennwerte für das vorliegende Datenset überhaupt Sinn macht.

Für die Einzelfallanalysen der Versuchspersonen sind Berechnungen von bspw. *RMSD* nicht möglich, da diese mit Hilfe der Versuchspersonenmittelwerte, der Modellmittelwerte und der Standardfehler dieser Mittelwerte durchgeführt werden müssen. Diese Berechnung findet bei Einzelfallanalysen allerdings nicht statt. Eine solche Mittelwertberechnung müsste über die Messzeitpunkte erfolgen, um damit einen mittleren zeitlichen Verlauf über alle Versuchspersonen zu erhalten. Dies ist allerdings auf Grund der fehlenden Standardisierung der Szenarien nicht als sinnvoll zu erachten, da durch das relativ zufällige Erscheinen der Flugzeuge innerhalb eines Szenarios die Höhe der Belastung zu den einzelnen Messzeitpunkten zwischen den Versuchspersonen variiert (vgl. Kapitel 6.1.2). Aus diesen Gründen kann die Messung der mittleren quadratischen Abweichung der exakten Lage von Datenpunkten zumindest für die zeitlichen Verläufe der Daten nicht erfolgen.

Für die Berechnung von Abweichungsmaßen der mittleren Beanspruchungshöhe innerhalb der vier Szenarien ergeben sich weitere Probleme: Zuallererst besteht das Problem der Skalierung der abhängigen Variablen (Schunn & Wallach, 2005): Im vorliegenden Experiment kann das Skalenniveau des Messinstrumentes (RSME-Fragebogen) nur als ordinal eingestuft werden. In diesem Fall sind Mittelwerte quantitativer Abweichungen von der exakten Lage des Datenpunktes nicht aussagekräftig, da die Bedeutung einer bestimmten Abweichung über die Skala hinweg variiert (die Skala ist nicht äquidistant). Bereits der Vergleich zweier Versuchspersonen würde starke Abweichungen ergeben, die auch jeweils beim Vergleich mit dem Modell auftreten würden. Diese Abweichungen sind allerdings nicht auf die Modellgüte zurückzuführen, sondern hängen mit der Subjektivität der Messskala zusammen.

Zusätzlich besteht das Problem, dass die mentale Beanspruchung über eine relativ arbiträre Skala gemessen wird, deren Messwerte sozusagen willkürlich festgelegt wurden. Es ist äußerst schwierig mit einem verbalen Anker zu beschreiben, welche Bedeutung ein Wert von „50“ auf

der RSME-Skala besitzt. Darüber hinaus wäre die absolute Position der Punkte in Bezug auf das Modell willkürlich, da zwei verschiedene Skalen verwendet werden (die RSME Skala von 0 bis 150 mit Ankerpunkten, die Modellskala von 0 bis 10 für die Chunks, die im Arbeitsgedächtnis prinzipiell vorhanden sein können). Durch eine Umformung der beiden Skalen in Prozentwerte kann zwar die Umwandlung in gleiche Einheiten erreicht werden, die Bedeutung der einzelnen Werte bleibt dennoch subjektiv und nicht äquidistant. In beiden Fällen bezeichnen Schunn und Wallach (2005) die Messung der exakten Lage von Datenpunkte als nicht informativ. Ebenso ist die festgelegte Maximalkapazität des Arbeitsgedächtnisses des Modells von zehn Chunks relativ willkürlich. Es kann angenommen werden, dass auch dieser Wert schwankt, wobei die prozentuale Beanspruchungshöhe vom Maximalwert abhängig ist, der festgelegt wurde.

Bei der Erhebung mentaler Beanspruchung ist daher die absolute Position eines Messwertes nicht so entscheidend wie der Trend der einzelnen Messwerte. Selbst bei absolut standardisierten Versuchsbedingungen und identischen Aufgaben und Belastungen können Versuchspersonen unterschiedliche Einschätzungen ihrer mentalen Beanspruchung abgeben, was von einem Modell, welches die individuellen Unterschiede der Personen nicht oder nur begrenzt betrachtet, nicht abgebildet werden kann.

Um dennoch eine **Einschätzung der Unterschiede zwischen Modell- und Realdaten** abgeben zu können, soll die absolute Abweichung (absolute deviation; AD) und mittlere absolute Abweichung (mean absolute deviation; MAD oder mean absolute error MAE) der mittleren Beanspruchungshöhe innerhalb jedes Szenarios angegeben werden. Diese Werte sind allerdings unter Beachtung der geschilderten Einschränkungen zu beurteilen. Daher wird auch keine Hypothese formuliert, sondern nur ein **exploratives Vorgehen** gewählt.

Exploratives Vorgehen: Berechnung der absoluten (AD) und mittleren absoluten Abweichung (MAD) der mittleren Beanspruchungshöhe des Modells von den experimentellen Daten (RSME und NASA-TLX) innerhalb jedes Szenarios.

Allgemein ist festzuhalten, dass bei der Erhebung mentaler Beanspruchung der relative Trend zwischen mehreren Versuchsbedingungen als sehr viel bedeutungsvoller einzuschätzen ist als die exakte Lage der Datenpunkte. Das bedeutet konkret, dass vor allem die qualitative Aussage, ob Szenario 1 (oder die Belastung zum Zeitpunkt 1) als höher beanspruchend empfunden wird als Szenario 2 (oder die Belastung zum Zeitpunkt 2) oder ob in Szenario 1 bereits die Belastungsgrenzen erreicht werden, von Bedeutung ist. Weniger informativ und auch schwer zu interpretieren ist dagegen die Aussage, dass die Belastung im Szenario 1 mit 20 %, in Szenario 2 aber nur mit 15 % eingeschätzt wurde, da aufgrund der subjektiven Natur der Messmethode nicht eindeutig festzulegen ist, was ein Beanspruchungswert von 20 % für eine Versuchsperson bedeutet.

Als Ergänzung zu den rein quantitativen Dimensionen beschreiben Schunn und Wallach (2005) auch die Fähigkeit des Modells auf neue Datensätze zu generalisieren. Diese Fähigkeit kann ermittelt werden, indem exakt dasselbe Modell (ohne Anpassung von Parametern) auf andere Datensätze (z.B. andere Stichproben oder andere Szenarien) angewendet wird. So kann im vorliegenden Experiment durch die Anwendung zweier Verkehrsszenarien für die beiden Belastungsstufen eine **Generalisierbarkeit des Modells** auf verschiedene Szenarien überprüft werden. Eine Überprüfung der Generalisierbarkeit auf eine weitere Stichprobe konnte aufgrund der fehlenden Ressourcen nicht stattfinden, sollte aber für folgende Forschung in Betracht gezogen werden. Ebenso sind die relative Komplexität und Falsifizierbarkeit des Modells Kriterien, welche in die Goodness-of-Fit Betrachtung einbezogen werden sollte.

Als weiteres, sozusagen hybrides Maß (qualitativ und quantitativ) für die Modellgüte kann **die relative Häufigkeit von Fehlern der Versuchspersonen in vom Modell replizierten Beanspruchungsgipfeln** herangezogen werden. Hierbei soll als **explorative Untersuchung** überprüft werden, ob ein aufgetretener Fehler in einem lokalen Maximum liegt, bzw. innerhalb eines lokalen Maximums mit einer Toleranz von ± 1 Minute. Zusätzlich zu diesem Kriterium kann außerdem definiert werden, dass, sollte der Modellwert fünf oder höher erreichen (>5 Chunks im Arbeitsgedächtnis), von einer erheblichen Beanspruchung ausgegangen werden kann, die auch Fehler wahrscheinlicher machen sollte.

Exploratives Vorgehen: Berechnung der relativen Häufigkeit, mit der Fehler von Versuchspersonen zu Zeitpunkten auftreten, in denen das Modell eine hohe Beanspruchung vorhersagt.

Alle Hypothesen und Untersuchungen zur Testung der Modellgüte sind in Tabelle 6 zusammengefasst.

Tabelle 6: Übersicht der Hypothesen zur Überprüfung der Modellgüte

	Nr.	Hypothese
Modelldaten	Maße der Modellgüte	
	7	Die mittleren Beanspruchungswerte des Modells sind bei hohem Verkehrsaufkommen signifikant höher als bei niedrigem. SZN 1, 3 > SZN 2, 4; SZN 1 = SZN 3; SZN 2 = SZN 4
	8	Für den Vergleich der mittleren Beanspruchungshöhe über die 4 Szenarien von Modell, subjektiven Maßen und Leistungsmaß besteht eine starke Korrelation (> 0,6).
	9	Es besteht ein positiver Zusammenhang zwischen den

		Beanspruchungsratings über die Zeit erfasst mit dem RSME und den Modelldaten.
	10	Es besteht ein negativer Zusammenhang zwischen den Beanspruchungsdaten über die Zeit erfasst durch das EKG (in ms IBI) und den Modelldaten.
	explorativ	Berechnung der absoluten und mittleren absoluten Abweichung (MAD) der mittleren Beanspruchungshöhe des Modells von den Realdaten (RSME und NASA-TLX)
	explorativ	Berechnung der relativen Häufigkeit, mit der Fehler von Versuchspersonen zu Zeitpunkten auftreten, in denen das Modell eine hohe Beanspruchung vorhersagt.

6.1.5 VERSUCHSABLAUF

Nach einer kurzen Begrüßung durch den Versuchsleiter wurden die Probanden zunächst allgemein über den Versuchsablauf und den Zweck der Studie informiert. Anschließend füllten die Versuchspersonen den demografischen Fragebogen, sowie die Einverständniserklärung aus.

Es folgte die Instruktion des d2-R Tests, sowie im Anschluss daran seine Bearbeitung. Dabei wurden die Versuchspersonen alle 20 Sekunden durch einen Piep-Ton aufgefordert, mit der nächsten Zeile zu beginnen. Im Anschluss an den d2-R Test erfolgte eine allgemeine Instruktion zur Flugverkehrskontrolle. Hierbei erhielten die Versuchspersonen Informationen über die generellen Aufgaben von Towerlotsen, sowie den Aufbau des Flughafens und der für Sie wichtigen Bereiche. Außerdem wurden der Ablauf der Funkkommunikation und die Verwendung der Phraseologie erläutert. Da, wie bereits erwähnt, die Verbindung der Simulationssoftware mit einer Spracherkennungssoftware nicht möglich war, wurde für die Spracherkennung ein Wizard-of-Oz Design verfolgt. Hierbei erhielt nur der Versuchsleiter die Anfragen der Flugzeuge aus der Simulation durch die vordefinierten Standardanfragen (in Englisch) und leitete diese per simulierter Funkkommunikation umgehend an die Versuchspersonen weiter (in Deutsch). Diese antworteten ebenfalls mündlich per Funkkommunikation mit der jeweiligen Freigabe, die dann vom Versuchsleiter per Tastenkürzel in den richtigen Text-Befehl für die Simulation übersetzt wurde. Hierbei bestand zusätzlich der Vorteil, dass die Versuchspersonen keine englischen Kommandos verwenden mussten, sondern, da es sich um Laien handelte, in Anlehnung an die englischen, auch in der Realität verwendeten Kommandos, selbst festgelegte deutsche Anfragen und Freigaben-Kommandos definiert werden konnten. Die Funkkommunikation erfolgte dabei mit Hilfe der Voice-over-IP Kommunikations-Software „Team Speak“, wobei die Versuchspersonen immer, wenn sie sprechen wollten, eine

markierte Taste auf einem Nummernpad gedrückt halten mussten (Push-to-Talk), was die Aktivierung des Mikrofons bei echten Lotsen abbildete. Es folgte die Instruktion der Hauptaufgabe, sowie die Beschreibung jeder einzelnen Anfrage und Freigabe. Diese umfasste die Erklärung aller drei auftretenden Anfragen von Flugzeugen („Fertig zum Rollen“, „Fertig zum Abflug“, „Bin bei Ihnen“ (Anfrage zum Landen)), sowie die dazugehörigen Freigaben („Rollen zur Piste“, „Startfreigabe“, „Landefreigabe“ und „Rollen zum Terminal“). Für das „Rollen zum Terminal“ wurde dabei keine spezifische Anfrage durch die Flugzeuge gegeben. Vielmehr mussten die Versuchspersonen selbst erkennen, wann ein Flugzeug von der Start-/Landebahn abgerollt und bereit war, eine weitere Freigabe zu erhalten, so wie es auch im Modell umgesetzt war. Als Gedankenstütze stand den Versuchspersonen den ganzen Versuch über eine Übersicht der Anfragen und der dazugehörigen Freigaben zur Verfügung.

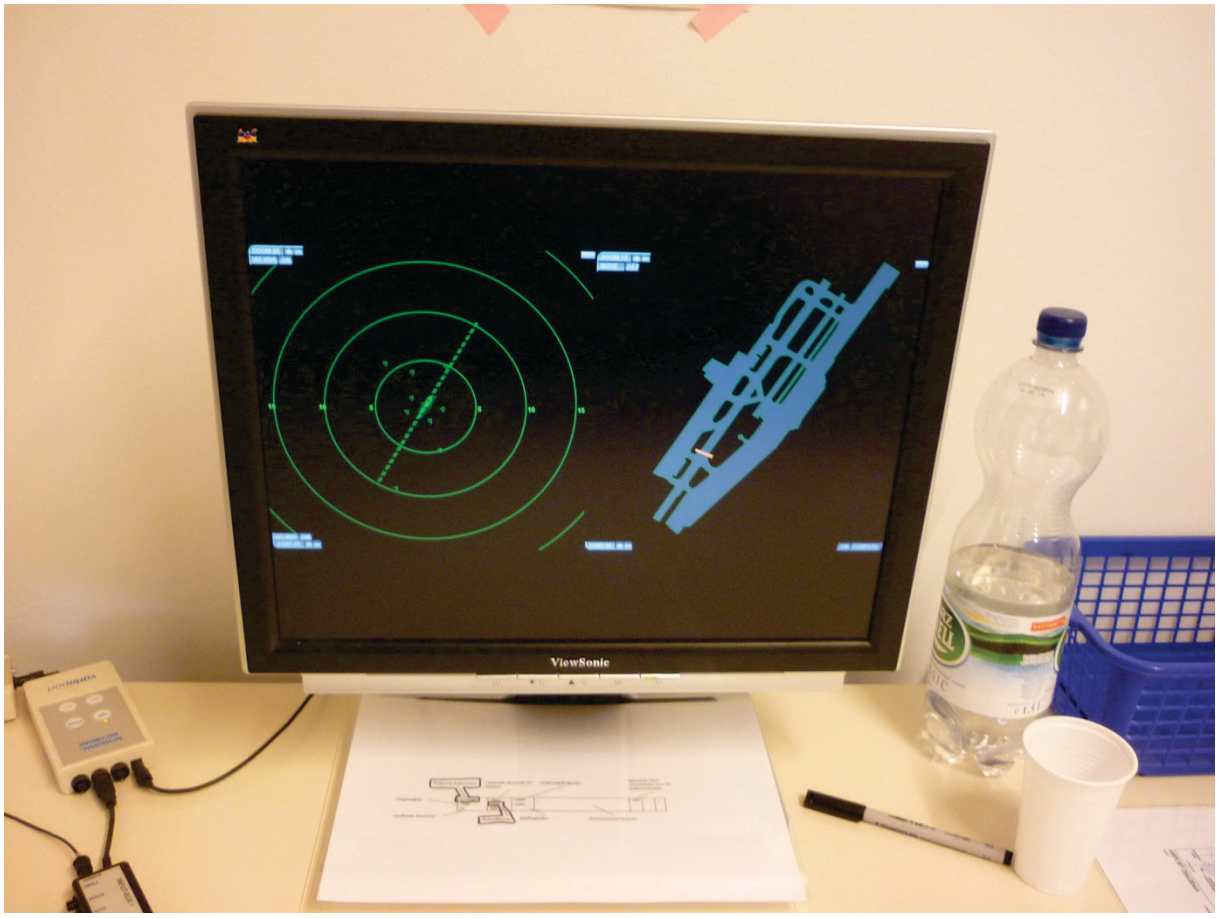

Abbildung 43: Arbeitsplatz der Versuchspersonen mit den zur Verfügung stehenden Radarbildschirmen (links Luftradar, rechts Bodenradar)

Nachdem die Hauptaufgabe instruiert war, erfolgte eine Beschreibung der Hilfsmittel. Hierbei standen einerseits die von der Simulation zur Verfügung gestellten Radarbildschirme für Boden- und Luftverkehr (vgl. Abbildung 43), andererseits nach realem Vorbild selbst erstellte Flugstreifen in einer Stripbay zur Verfügung (vgl. Abbildung 44). Auf dem Luftradar waren dabei die Flugbewegungen in einiger Entfernung um den Flughafen zu erkennen. Das Bodenradar stellte eine Karte des Flughafens mit den jeweiligen Positionen der rollenden Luftfahrzeuge dar.

Die Flugstreifen wurden in der Stripbay vor jedem Szenario vorsortiert und standen den Versuchspersonen zum Vermerk von Notizen zu jedem Flugzeug zur Verfügung. Es wurde entschieden, die Flugstreifen in der Stripbay vorzusortieren, da sonst die Beanspruchung der Versuchspersonen zu hoch gewesen wäre.

Im Anschluss an die ausführliche Instruktion wurden die EKG-Elektroden gesetzt und die Versuchspersonen aufgefordert die von ihnen auszuführende Aufgabe nochmals kurz mit eigenen Worten zusammenzufassen. Anschließend begann eine 5-minütige Trainingsphase, in der ein Verkehrsszenario bearbeitet wurde, in dem alle Anfragen und Freigaben einmal auftraten. Um die Versuchsbedingungen möglichst genau zu trainieren, wurden die Versuchspersonen hier bereits jede Minute durch einen Piep-Ton aufgefordert, mittels RSME-Fragebogen eine Einschätzung ihrer momentanen mentalen Beanspruchung abzugeben, sowie am Ende des Szenarios mittels NASA-TLX eine Einschätzung ihrer mentalen Beanspruchung über das gesamte Szenario abzugeben.

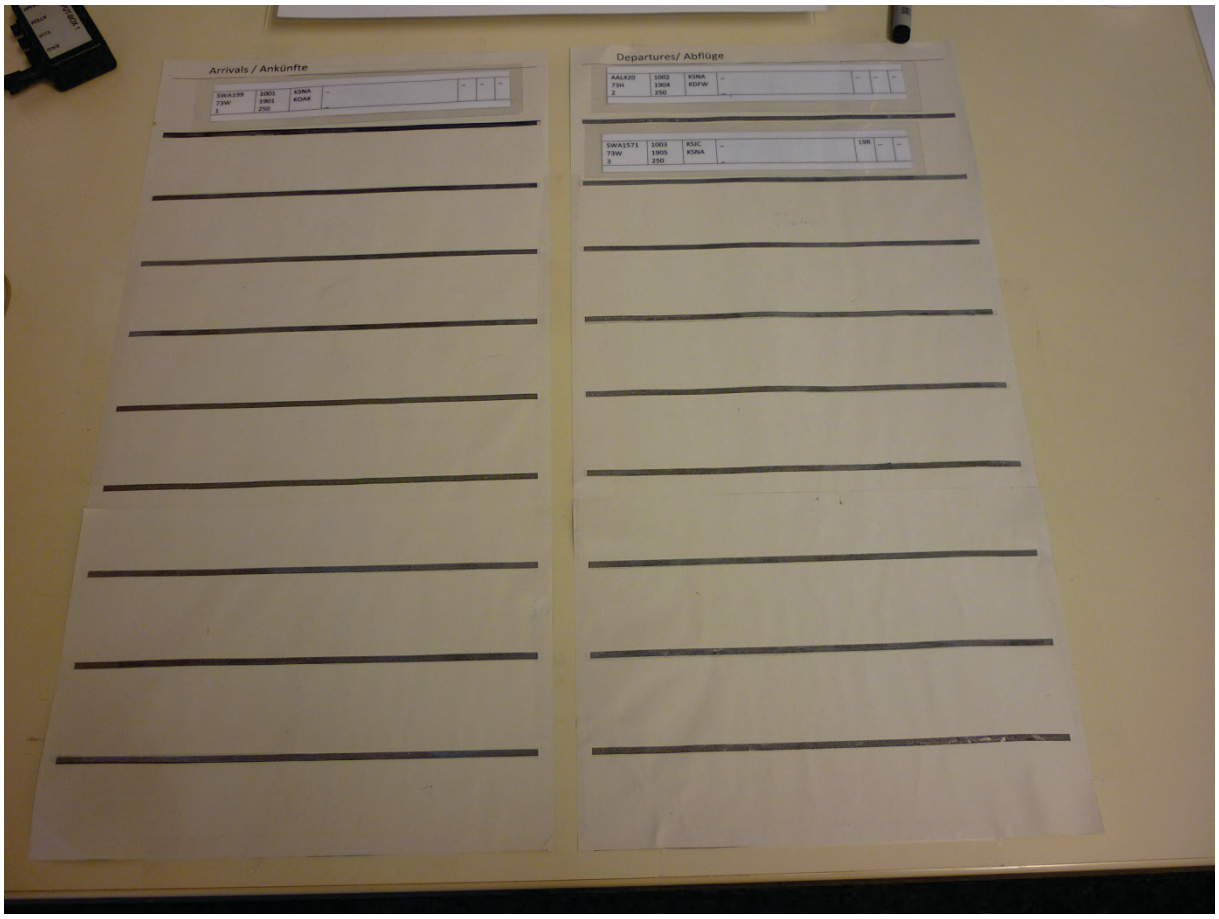

Abbildung 44: Die den Versuchspersonen zur Verfügung stehende Stripbay mit einigen Flugstreifen

Im Anschluss startete der eigentliche Versuch. Hierbei hatten alle Versuchspersonen alle vier 15-minütigen Verkehrsszenarien zu bearbeiten. Vor jedem Szenario wurde außerdem eine 5-minütige Baselinemessung des EKG durchgeführt. Der gesamte Versuch dauerte zwei Stunden. Innerhalb der Szenarien erhielten die Versuchspersonen jeweils Anfragen von Flugzeugen (vom

Versuchsleiter) auf die sie die richtigen Freigaben erteilen mussten, um den flüssigen und sicheren Verkehr am Flughafen zu gewährleisten. Dabei wurden die Versuchspersonen ausdrücklich instruiert, dass die Abfertigung von Landeanfragen, sowie das Abrollen von der Piste mit Priorität zu behandeln ist und dass es daher möglich ist, Flugzeuge an der Parkposition oder am Haltepunkt der Start-/Landebahn warten zu lassen. Nichtsdestotrotz sollten die Versuchspersonen die wartenden Flugzeuge flüssig abwickeln, falls die Verkehrssituation es zulässt. Im Hinblick auf eine sichere Abwicklung der Flüge, wurde definiert, dass sich zu keiner Zeit zwei Flugzeug auf der Start-/Landebahn aufhalten dürfen, dass aber durchaus zwei Freigaben für diese vorliegen können, sofern das eine Flugzeug die Piste verlassen hat, bevor das andere auf diese aufsetzt oder aufrollt. Für den Bodenverkehr galt ebenso, dass die Rollfreigaben so erteilt werden sollten, dass es nicht zu Kollisionen kommt.

Nachdem alle vier Szenarien bearbeitet waren, wurden die EKG-Elektroden entfernt und die Versuchspersonen wurden verabschiedet. Für die Teilnahme erhielten sie eine Aufwandentschädigung von 20 €.

6.2 Ergebnisse

Die Darstellung der Ergebnisse erfolgt in der Reihenfolge der beschriebenen Hypothesen 1 – 10.

6.2.1 Zusammenhang d2-R Ergebnis und Fehlerzahl bzw. Höhe der Beanspruchungsratings (Hypothese 1)

Hypothese (1): Versuchspersonen mit besserem Abschneiden im d2-R-Test, machen auch in der Fluglotsenaufgabe insgesamt weniger Fehler und schätzen die mentale Beanspruchung (RSME und NASA-TLX) insgesamt als geringer ein.

Mit dieser Analyse sollte festgestellt werden, ob das Abschneiden der Versuchsperson im d2-R-Test einen Einfluss auf die Anzahl der gemachten Fehler und der abgegebenen subjektiven Beanspruchungseinschätzung hat. Hierzu wurden anhand der im d2-R-Test vorgegebenen Einteilung folgende Gruppen anhand der Gesamtpunktzahl definiert: hoher d2-R-Wert (KL ≥ 106), durchschnittlicher d2-R-Wert (94 < KL < 106) und geringer d2-R-Wert (KL ≤ 94). Somit wurden drei Gruppen erhalten, wobei acht Versuchspersonen in die Gruppe „hoher d2-R-Wert“ eingeordnet werden konnten, weitere sechs Versuchspersonen in die Gruppe „geringer d2-R-Wert“ und zehn Versuchspersonen in die Gruppe „durchschnittlicher d2-R-Wert“. Die Abbildung 45 lässt zumindest deskriptiv einen Unterschied zwischen den einzelnen Gruppen und der mittleren Fehlerrate erkennen. Aufgrund der Tatsache, dass die Fehlerzahlen nicht normalverteilt sind (Kolmogorov-Smirnoff-Test auf Normalverteilung der Daten liefert für die Reihen der Szenarien 2 und 4 signifikante Ergebnisse), wurde eine Friedmann-ANOVA durchgeführt. Diese überschritt knapp das Signifikanzniveau von $\alpha = 0{,}05$ ($\chi^2 = 6{,}333$; df 2; p =

0,052), so dass kein signifikanter Zusammenhang zwischen den Fehlerzahlen und dem Abschneiden im d2-R-Test angenommen werden kann.

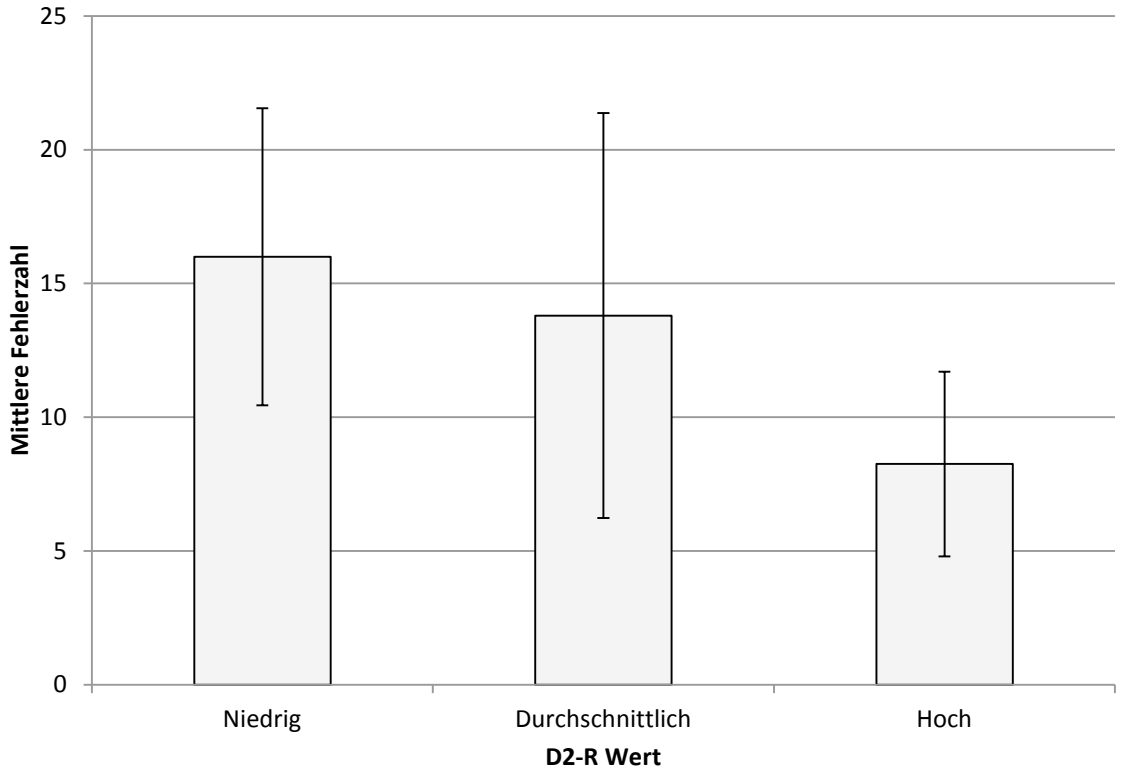

Abbildung 45: Mittlere Fehlerzahl der drei Gruppen nach ihrem Abschneiden im d2-R-Test

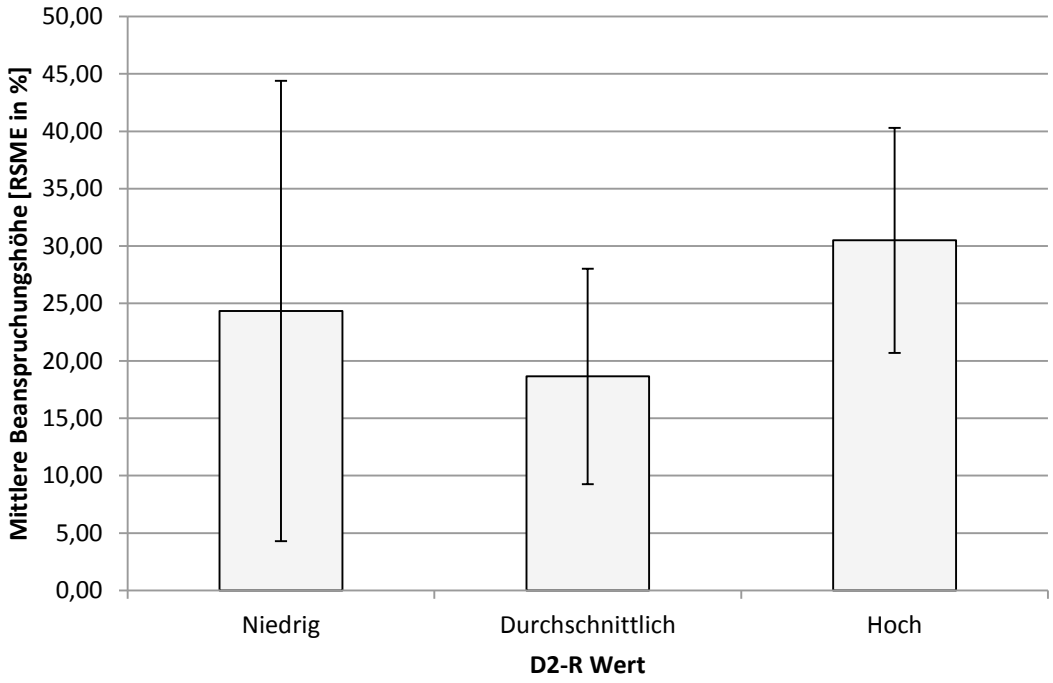

Abbildung 46: Mittlere Beanspruchungshöhe (laut RSME) der drei Gruppen nach d2-R-Ergebiniss

Weiterhin wurde untersucht, ob ein Zusammenhang zwischen dem Abschneiden im d2-R-Test und der subjektiven Beanspruchungshöhe erfasst mit dem RSME und dem NASA-TLX (Sub-Skala geistige Anstrengung (g. A.)) besteht. Laut Abbildung 46 und Abbildung 47 ist hier deskriptiv zu erkennen, dass diejenigen Versuchspersonen mit einem hohen d2-R-Wert die Beanspruchung im Mittel am höchsten einschätzten, die mit durchschnittlichem Abschneiden am niedrigsten. Das gleiche Bild ergibt sich bei den Bewertungen mittels NASA-TLX (g. A.)[65]. Aufgrund des Skalenniveaus der Daten (nur Ordinalskalenniveau kann angenommen werden) wurde auch hier eine Friedmann-ANOVA durchgeführt.

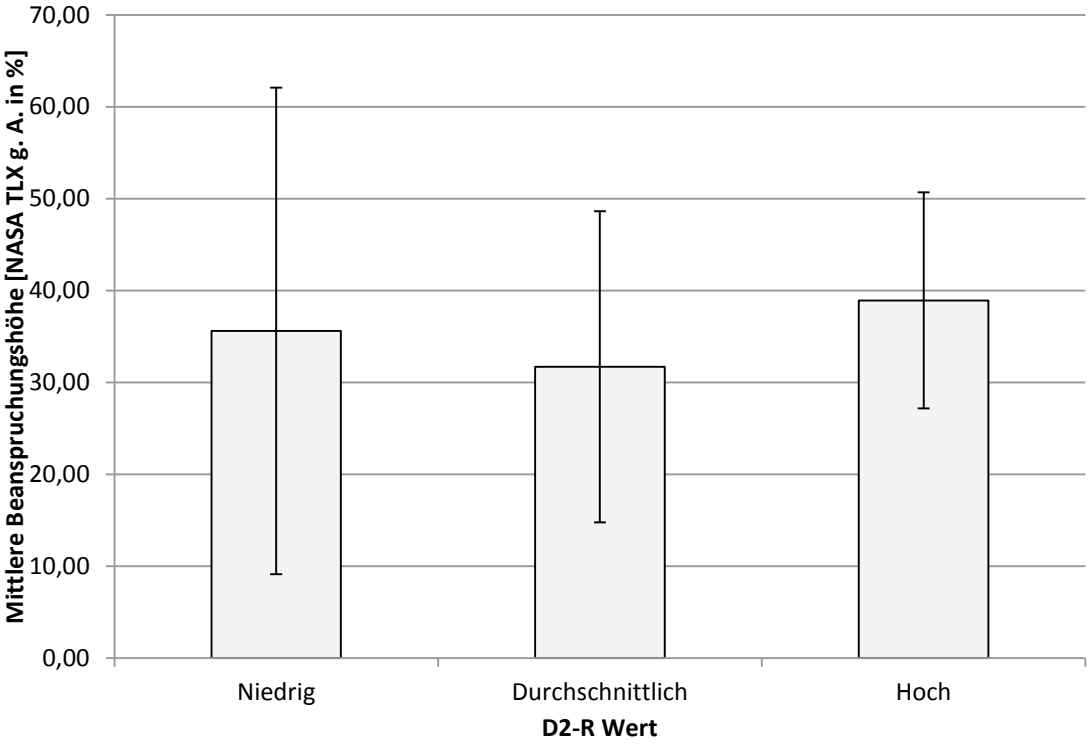

Abbildung 47: Mittlere Beanspruchungshöhe (laut NASA-TLX g. A.) der drei Gruppen nach d2-R-Ergebniss

Diese gelangte für beide Maße zu keinen signifikanten Ergebnissen (RSME: $\chi^2 < 1{,}000$; df 2; $p = 0{,}704$; NASA-TLX g. A.: $\chi^2 < 0{,}333$, df 2; $p = 0{,}956$). Für die weitere Auswertung der subjektiven Maße sowie des Leistungsmaßes kann daher ein Einfluss des Abschneidens im d2-R-Test und damit des Effizienz-Geschwindigkeits-Tradeoff auf die Beanspruchungsbewertungen ausgeschlossen werden.

Tabelle 7: Ergebnisse der Testung der Hypothese 1

Hypothese	Getätigte Annahme	Ergebnis
Hypothese 1a	Mit steigendem d2-R-Wert werden weniger Fehler gemacht.	Nicht bestätigt ✗

[65] Die absoluten Skalenwerte der subjektiven Maße wurden zur besseren Vergleichbarkeit jeweils in Prozentwerte überführt.

Hypothese 1b	Mit steigendem d2-R-Wert wird die Beanspruchung als geringer eingeschätzt.	Nicht bestätigt ✖

6.2.2 Bewertung der Sensitivität der Herzrate (Hypothese 2)

Hypothese (2): Die Herzrate, gemessen in Millisekunden Dauer der Inter-Beat-Intervalle, der beiden Experimentalbedingungen (hohes vs. niedriges Verkehrsaufkommen) ist signifikant geringer als die der Baseline-Bedingung. Die mittleren IBI der Szenarien mit hohem Verkehrsaufkommen (1 und 3) sind geringer, als die mittleren IBI der Szenarien mit geringem Verkehrsaufkommen (2 und 4).

Um festzustellen, ob im vorliegenden Experiment eine genügend hohe Sensitivität der Herzrate zur Unterscheidung zweier Experimentalbedingungen vorlag, wurden die Experimental-bedingungen mit der Baseline-Bedingung verglichen. Die Stufe der Belastung diente hierbei als Faktorstufe. Daher wurden die drei Stufen hohe Belastung (Szenarien 1 und 3), niedrige Belastung (Szenarien 2 und 4) und Baseline definiert. Deskriptiv kann hierbei ein leichter Anstieg der mittleren Dauer der Inter-Beat-Intervalle von der Bedingung mit hoher Belastung mit $\overline{IBI}$ = 747,65 ms (s = 61,25 ms), über niedrige Belastung mit $\overline{IBI}$ = 752,96 ms (s = 58,95 ms) bis zur Baseline mit $\overline{IBI}$ = 771,77 ms (s = 62,76 ms) festgestellt werden (vgl. Abbildung 48).

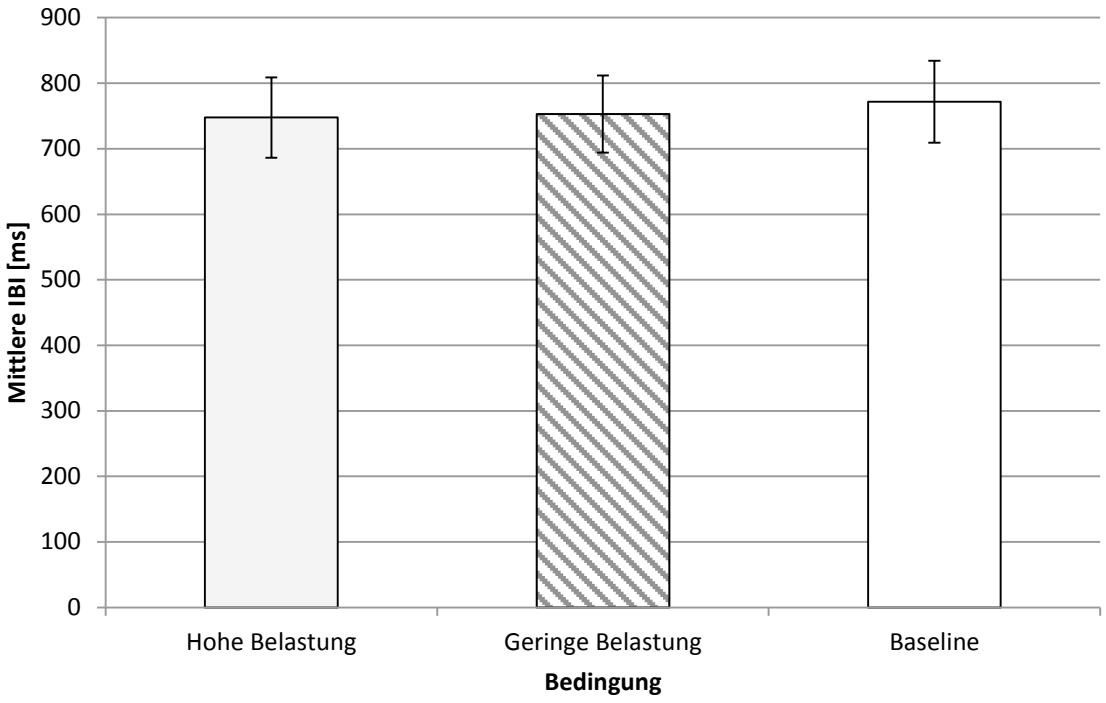

Abbildung 48: Mittlere IBI (in ms) der beiden Experimentalbedingungen (hohe und geringe Belastung) und der Baselinebedingung

Da für die EKG-Daten (Herzrate in IBI) Intervallskalierung, sowie Normalverteilung (Kolmogorov-Smirnoff-Test nicht signifikant) und Varianzhomogenität (Lavenes Test nicht

signifikant) gegeben war, wurde zur Signifikanztestung eine Messwiederholungs-ANOVA mit geplanten Kontrasten für die drei Bedingungen durchgeführt. Die Hauptanalyse ergab, dass die Länge der Inter-Beat-Intervalle hoch signifikant von der Belastung beeinflusst wird; $F(2, 94) = 15{,}4$, $p < 0{,}001$. Eine gesamte Effektstärke von 0,299 (partielles η^2) zeigt, dass etwa 30 % der Varianz auf die veränderte Belastung zurückgeführt werden kann. Die paarweisen Vergleiche ergaben, dass die Inter-Beat-Intervalle in der Baseline-Bedingung signifikant länger sind als in beiden Experimentalbedingungen (hohe Belastung Mittelwertsdifferenz 0,252; $p < 0{,}001$; CI(95%) 0,117-0,387; geringe Belastung Mittelwertsdifferenz 0,199; $p < 0{,}001$; CI(95%) 0,089-0,309). Die Länge der Inter-Beat-Intervalle der beiden Experimentalbedingungen unterscheidet sich allerdings nur minimal (Mittelwertsdifferenz 0,053; $p = 0{,}691$). So können die Hypothesen 2a und 2b nicht widerlegt werden, während die Hypothese 2c abgelehnt wird.

Es kann somit davon ausgegangen werden, dass die Herzrate nicht in der Lage ist, feine Beanspruchungsunterschiede aufzudecken, wodurch sie aus den weiteren Betrachtungen ausgeschlossen werden muss, da kein Informationsmehrwert zu erwarten ist.

Tabelle 8: Ergebnisse der Testung der Hypothese 2

Hypothese	Getätigte Annahme	Ergebnis
Hypothese 2a	Die mittleren IBI der Szenarien mit hohem Verkehrsaufkommen (1 und 3) sind geringer als die der Baseline Bedingung (Mittelwertsdifferenz 0,252; $p < 0{,}001$; CI(95%) 0,117-0,387).	Bestätigt ✓
Hypothese 2b	Die mittleren IBI der Szenarien mit geringem Verkehrsaufkommen (2 und 4) sind geringer als die der Baseline Bedingung (Mittelwertsdifferenz 0,199; $p < 0{,}001$; CI(95%) 0,089-0,309).	Bestätigt ✓
Hypothese 2c	Die mittleren IBI der Szenarien mit hohem Verkehrsaufkommen (1 und 3) sind geringer als die der Szenarien mit geringem Verkehrsaufkommen (2 und 4) (Mittelwertsdifferenz 0,053; $p = 0{,}691$).	Nicht bestätigt ✗

6.2.3 Bewertung der mittleren mentalen Beanspruchung mittels Herzrate (Hypothesen 3a- f)

Hypothese (3): Die Herzrate, gemessen in Millisekunden Dauer der Inter-Beat-Intervalle, ist bei hohem Verkehrsaufkommen signifikant geringer als bei niedrigem Verkehrsaufkommen. Dagegen unterscheidet sich die Herzrate von Szenarien mit gleich hohem Verkehrsaufkommen nicht.

Aufgrund der zur Überprüfung der Hypothese 2 durchgeführten Analysen kann durch die geringe Sensitivität der Herzrate keine Aussage bezüglich der Hypothese 3 gemacht werden. In

der Analyse zu Hypothese 2 zeigte sich bereits, dass keine signifikanten Unterschiede in der Herzrate innerhalb der beiden Experimentalbedingungen bestehen. Damit kann die Herzrate nicht als verlässlicher Indikator für Beanspruchungsunterschiede in den Experimentalbedingungen betrachtet und im Weiteren nicht als Maß zur Hypothesenprüfung eingesetzt werden.

6.2.4 Bewertung der mittleren mentalen Beanspruchung mittels RSME (Hypothesen 4a – f)

Hypothese (4): Die subjektiven Bewertungen der mentalen Beanspruchung, erhoben mit der RSME-Skala, sind bei hohem Verkehrsaufkommen signifikant höher als bei niedrigem. Die Bewertungen von Szenarien mit gleich hohem Verkehrsaufkommen unterscheiden sich nicht voneinander.

In den Szenarien 1 und 3 (hohes Verkehrsaufkommen) schätzten die Probanden die Beanspruchung im Mittel mit $Median_{SZN1}$ = 25,06 % (MD[66] = 10,30 %) respektive $Median_{SZN3}$ = 29,39 % (MD = 11,08 %) ein. In den Szenarien 2 und 4 wurde die Beanspruchung im Mittel mit $Median_{SZN2}$ = 16,52 % (MD = 10,15 %) bzw. $Median_{SZN4}$ = 17,29 % (MD = 9,85 %) eingeschätzt (vgl. Abbildung 49). Es ist deskriptiv also zu erkennen, dass die mentale Beanspruchung auch in Szenarien mit gleicher Belastung leicht unterschiedlich bewertet worden ist.

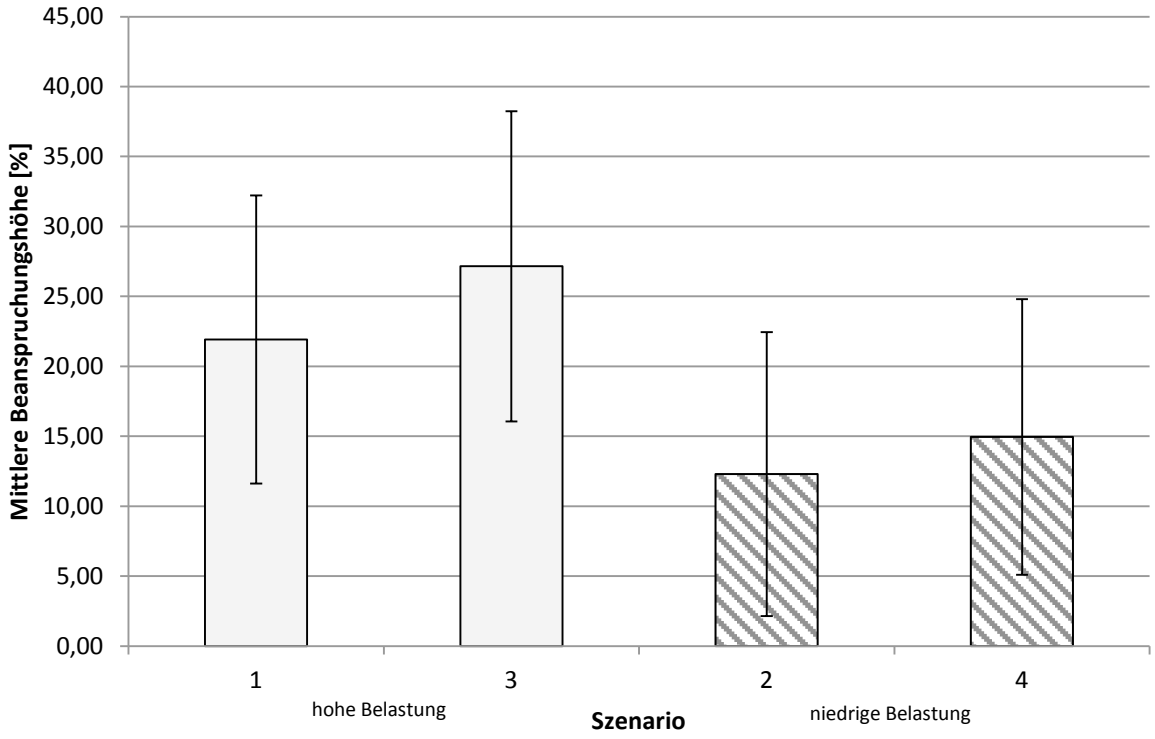

Abbildung 49: Mittlere Beanspruchungshöhe jedes Szenarios (in %) erfasst mit dem RSME

[66] MD: Mittlere Abweichung vom Median.

Aufgrund des Ordinalskalenniveaus wurde eine Friedmann-ANOVA durchgeführt, welche auf hoch signifikante Unterschiede zwischen den Bedingungen hindeutete ($\chi 2 = 33{,}734$, df 3, $p < 0{,}001$). Für die paarweisen Vergleiche der Szenarien untereinander wurden post-hoc Wilcoxon-Tests durchgeführt, wobei sowohl vier Alternativhypothesen, als auch zwei Nullhypothesen als Forschungshypothesen formuliert worden sind. Dies ist beim Setzen des α-Niveaus zu berücksichtigen (vgl. Kapitel 6.1.4). Für die Testung der Alternativhypothesen ist bei der Berechnung von sechs post-hoc Wilcoxon-Tests eine Bonferroni-Korrektur durchzuführen, um eine Inflation des α-Fehlers zu vermeiden. Für die vier Alternativhypothesen gilt bei einseitiger Testung damit ein α-Niveau von $\alpha = \frac{0{,}05}{6} = 0{,}0083$. Für die Bestätigung der Nullhypothesen wird hingegen ein konservatives Vorgehen gewählt, wobei das α-Niveau auf $\alpha = 0{,}20$ gesetzt und keine Bonferroni-Korrektur vorgenommen wird, um eine hohe Teststärke zu erhalten und somit die Annahme der Nullhypothese zu erschweren. Für die vier Alternativhypothesen ergaben sich folgende Ergebnisse:

Tabelle 9: Ergebnisse der Testung der Alternativhypothesen von Hypothese 4

Hypothese	Getätigte Annahme	Ergebnis
Hypothese 4a	Die RSME-Ratings von SZN 1 sind signifikant höher als von SZN 2 ($T = 25$, $p < 0{,}001$, $r = 0{,}496$)	Bestätigt ✓
Hypothese 4b	Die RSME-Ratings von SZN 1 sind signifikant höher als von SZN 4 ($T = 30$, $p < 0{,}001$, $r = -0{,}495$)	Bestätigt ✓
Hypothese 4c	Die RSME-Ratings von SZN 3 sind signifikant höher als von SZN 2 ($T = 13$, $p < 0{,}001$, $r = -0{,}565$)	Bestätigt ✓
Hypothese 4d	Die RSME-Ratings von SZN 3 sind signifikant höher als von SZN 4 ($T = 15$, $p < 0{,}001$, $r = -0{,}557$)	Bestätigt ✓

Die vier Alternativhypothesen (4a – d) gelten damit.

Für die Nullhypothesen 4e – f ergaben sich folgende Ergebnisse:

Tabelle 10: Ergebnisse der Testung der Nullhypothesen von Hypothese 4

Hypothese	Getätigte Annahme	Ergebnis
Hypothese 4e	Die RSME-Werte von SZN 1 und SZN 3 unterscheiden sich nicht ($T = 103{,}5$, p (2-seitig) $= 0{,}190$, $r = -0{,}192$)	Nicht bestätigt ✗
Hypothese 4f	Die RSME-Werte von SZN 2 und SZN 4 unterscheiden sich nicht ($T = 99$, p (2-seitig) $= 0{,}384$, $r = -0{,}192$)	Bestätigt ✓

Die Hypothese 4e unterschreitet knapp das für die Testung der Nullhypothese gesetzte Signifikanzniveau von 0,20 (p = 0,19), sodass hier davon ausgegangen werden kann, dass ein Unterschied in den Bewertungen besteht. Für die Nullhypothese 4f werden keine signifikanten Ergebnisse erhalten und damit gilt diese ebenfalls.

6.2.5 Bewertung der mittleren mentalen Beanspruchung mittels NASA-TLX Dimension geistige Anstrengung (Hypothesen 5a – f)

Hypothese (5): Die subjektiven Bewertungen der mentalen Beanspruchung, erhoben mit dem NASA-TLX (Sub-Skala geistige Anstrengung), sind bei hohem Verkehrsaufkommen signifikant höher als bei niedrigem. Die Bewertungen von Szenarien mit gleich hohem Verkehrsaufkommen unterscheiden sich nicht voneinander.

In den Szenarien 1 und 3 (hohes Verkehrsaufkommen) schätzten die Probanden die Beanspruchung im Mittel mit $Median_{SZN1}$ = 63,54 % (MD = 20,24 %) respektive $Median_{SZN3}$ = 50,93 % (MD = 21,40 %) ein. In den Szenarien 2 und 4 wurde die Beanspruchung im Mittel mit $Median_{SZN2}$ = 34,51 % (MD = 19,38 %) bzw. $Median_{SZN4}$ = 31,47 % (MD = 20,41 %) eingeschätzt (vgl. Abbildung 50).

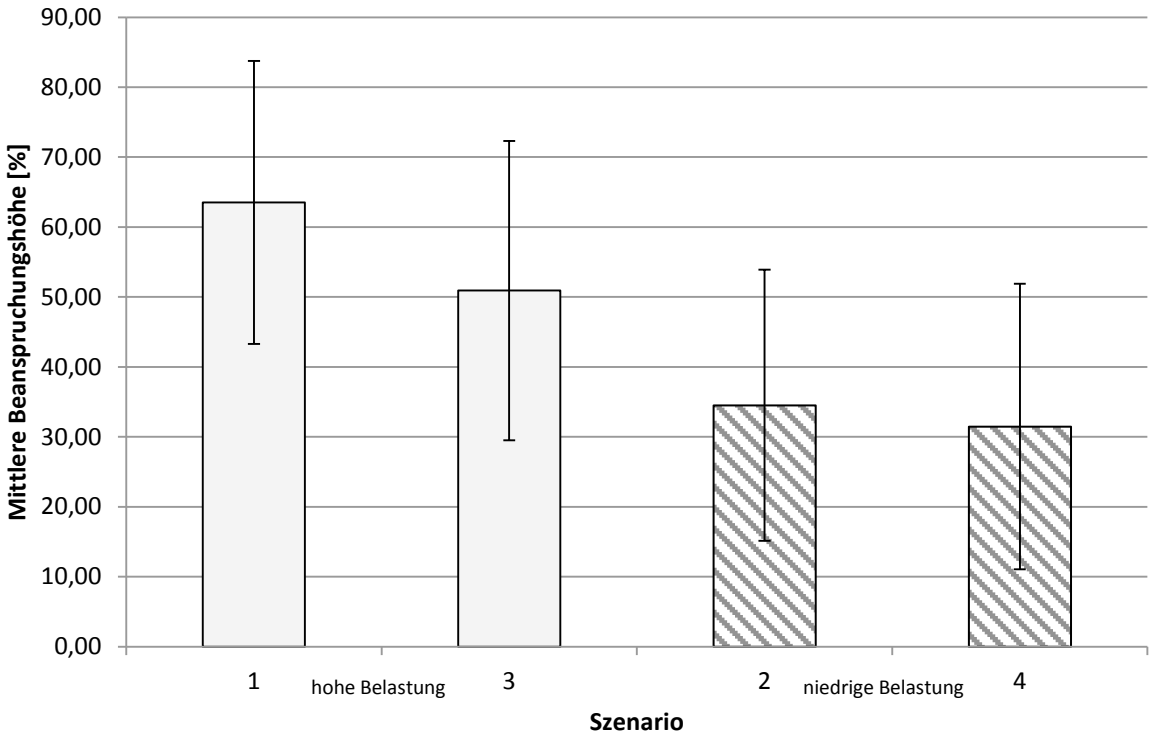

Abbildung 50: Mittleren Beanspruchungshöhe jedes Szenarios (in %) erfasst mit dem NASA-TLX (Dimension geistige Anstrengung)

Es ist deskriptiv also zu erkennen, dass die mentale Beanspruchung auch in Szenarien mit gleicher Belastung leicht unterschiedlich bewertet worden ist. Es ist ebenfalls zu erkennen, dass

die Beanspruchung mit dem NASA-TLX sehr viel höher (etwa doppelt so hoch[67]) eingeschätzt worden ist, als mit dem RSME (vgl. Hypothese 4, Kapitel 6.2.4). Ferner ist zu erkennen, dass im Vergleich zum RSME die Beanspruchungshöhe der einzelnen Szenarien unterschiedlich eingeschätzt worden ist. So ist mit dem RSME beispielsweise das Szenario 3 am höchsten bewertet worden, mit dem NASA-TLX hingegen das Szenario 1.

Eine Friedmann-ANOVA deutete auf hoch signifikante Unterschiede zwischen den Bedingungen hin ($\chi 2 = 29{,}038$, df 3, $p < 0{,}001$). Die Berechnung des α-Niveaus ergibt sich analog den Erläuterungen im vorangegangenen Kapitel für die Alternativhypothesen zu $\alpha = 0{,}0083$ und die Nullhypothesen zu $\alpha = 0{,}20$. Die Testung der vier Alternativhypothesen (2a – 2d) führen zu folgenden Ergebnissen:

Tabelle 11: Ergebnisse der Testung der Alternativhypothesen von Hypothese 5

Hypothese	Getätigte Annahme	Ergebnis
Hypothese 5a	Das NASA-TLX-Rating von SZN 1 ist signifikant höher als von SZN 2 ($T = 24$, $p < 0{,}001$, $r = -0{,}519$)	Bestätigt ✓
Hypothese 5b	Das NASA-TLX-Rating von SZN 1 ist signifikant höher als von SZN 4 ($T = 6$, $p < 0{,}001$, $r = -0{,}549$)	Bestätigt ✓
Hypothese 5c	Das NASA-TLX-Rating von SZN 3 ist signifikant höher als von SZN 2 ($T = 37$, $p = 0{,}001$, $r = -0{,}419$)	Bestätigt ✓
Hypothese 5d	Das NASA-TLX-Rating von SZN 3 ist signifikant höher als von SZN 4 ($T = 44$, $p = 0{,}001$, $r = -0{,}437$)	Bestätigt ✓

Die vier Alternativhypothesen 5a –d gelten damit.

Für beide Nullhypothesen 5e – f ergaben sich folgende Ergebnisse:

Tabelle 12: Ergebnisse der Testung der Nullhypothesen von Hypothese 5

Hypothese	Getätigte Annahme	Ergebnis
Hypothese 5e	Das NASA-TLX-Ratings von SZN 1 und SZN 3 unterscheiden sich nicht ($T = 114{,}5$, p (2-seitig) $= 0{,}320$, $r = -0{,}146$)	Bestätigt ✓
Hypothese 5f	Das NASA-TLX-Ratings von SZN 2 und SZN 4 unterscheiden sich nicht ($T = 134{,}5$, p (2-seitig) $= 0{,}923$, $r = -0{,}015$)	Bestätigt ✓

Es werden keine signifikanten Ergebnisse erhalten, womit diese weiterhin gelten.

[67] Aufgrund der Überführung der Skalenwerte des NASA-TLX und des RSME in Prozentwerte ist zwar ein Vergleich der Beanspruchungshöhe möglich, dieser ist allerdings unter Berücksichtigung der in Kapitel 6.1.4 gemachten Einschränkungen zu bewerten.

6.2.6 *Fehlerzahl (Hypothesen 6a – f)*

Hypothese (6): Die Fehlerzahlen sind bei hohem Verkehrsaufkommen signifikant höher als bei niedrigem. Die Fehlerzahlen von Szenarien mit gleich hohem Verkehrsaufkommen unterscheiden sich nicht voneinander.

Die Anzahl der Fehler wurde als Leistungsmaß erhoben. Deskriptiv ist zu erkennen, dass die Häufigkeit, mit der Fehler gemacht wurden, in den Szenarien mit hohem Verkehrsaufkommen sehr viel höher ist als bei geringem Verkehrsaufkommen (Abbildung 51).

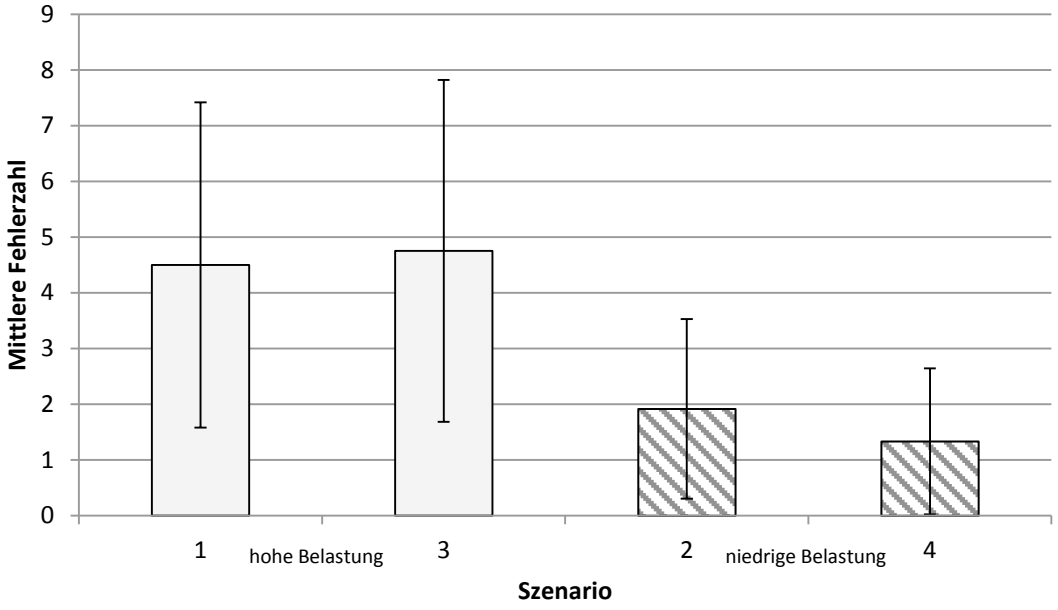

Abbildung 51: Mittlere Gesamtfehlerzahl pro Szenario

Es wurde überprüft, ob sich die Fehlerzahlen der einzelnen Szenarien signifikant voneinander unterscheiden. Laut Hypothesen 6a – f sollte dies bei den Vergleichen von Szenario 1 und 2, 1 und 4, 3 und 2 sowie 3 und 4 der Fall sein, bei den Szenarien 1 und 3 sowie 2 und 4 jeweils nicht. Aufgrund der Tatsache, dass die Fehlerzahlen nicht normalverteilt waren (Kolmogorov-Smirnoff-Test auf Normalverteilung der Daten lieferte für die Reihen der Szenarien 2 und 4 signifikante Ergebnisse), wurde zunächst eine Friedmann-ANOVA verwendet und mit post-hoc Wilcoxon Tests die einzelnen Fälle verglichen. Im Mittel machten die Versuchspersonen im Szenario 1 MW_{SZN1} = 4,54 Fehler (s = 2,89), im Szenario 2 MW_{SZN2} = 2,00 Fehler (s = 1,56), im Szenario 3 MW_{SZN3} = 4,63 Fehler (s = 3,16) und im Szenario 4 MW_{SZN4} = 1,25 Fehler (s = 1,26). Eine Friedmann-ANOVA lieferte hoch signifikante Ergebnisse ($\chi 2$ = 26,294, df 3, $p < 0{,}001$). Die Berechnung des α-Niveaus der post-hoc Wilcoxon-Tests ergibt sich analog den Erläuterungen im vorangegangenen Kapitel für die Alternativhypothesen zu α = 0,0083 und die Nullhypothesen zu α = 0,20. Die Testung der vier Alternativhypothesen (3a – d) führten zu signifikanten Ergebnissen:

Tabelle 13: Ergebnisse der Testung der Alternativhypothesen von Hypothese 6

Hypothese	Getätigte Annahme	Ergebnis
Hypothese 6a	Die mittlere Fehlerzahl von SZN 1 ist höher als von SZN 2 (T = 24,5, p < 0,001, r = -0,500)	Bestätigt ✓
Hypothese 6b	Die mittlere Fehlerzahl von SZN 1 ist höher als von SZN 4 (T = 10,5, p < 0,001, r = -0,547)	Bestätigt ✓
Hypothese 6c	Die mittlere Fehlerzahl von SZN 3 ist höher als von SZN 2 (T = 20, p < 0,001, r = -0,519)	Bestätigt ✓
Hypothese 6d	Die mittlere Fehlerzahl von SZN 3 ist höher als von SZN 4 (T = 1,5, p < 0,001, r = -0,530)	Bestätigt ✓

Somit gelten die Alternativhypothesen 6a – d.

Für beide Nullhypothesen 6e – f ergaben sich folgende Ergebnisse:

Tabelle 14: Ergebnisse der Testung der Nullhypothesen von Hypothese 6

Hypothese	Getätigte Annahme	Ergebnis
Hypothese 6e	Die mittlere Fehlerzahlen von SZN 1 und SZN 3 unterscheiden sich nicht (T = 113,5, p (2-seitig) = 0,953, r = -0,101)	Bestätigt ✓
Hypothese 6f	Die mittlere Fehlerzahlen von SZN 2 und SZN 4 unterscheiden sich nicht (T = 45,5, p (2-seitig) = 0,045, r = -0,292)	Nicht Bestätigt ✗

Für die Nullhypothese 6e wird kein signifikantes Ergebnis erhalten, sodass diese weiterhin gilt. Die Hypothese 6f unterschreitet das Signifikanzniveau von $\alpha = 0{,}20$; es kann hier davon ausgegangen werden, dass ein Unterschied in der mittleren Anzahl der Fehler besteht, sodass die Nullhypothese verworfen werden muss.

6.2.7 Mittlere Beanspruchungshöhe des Modells (Hypothese 7)

Hypothese (7): Die mittleren Beanspruchungswerte des Modells sind bei hohem Verkehrsaufkommen signifikant höher als bei niedrigem. Die mittleren Beanspruchungswerte des Modells von Szenarien mit gleich hohem Verkehrsaufkommen unterscheiden sich nicht voneinander.

Für die Modelldaten wurde das gleiche Bild erwartet, wie es auch in den subjektiven Bewertungen gezeigt werden konnte. Im Mittel ergab das Modell für die beiden Szenarien mit hoher Belastung höhere Werte von MW_{SZN1} = 33,92 % (s = 6,31 %) für Szenario 1 und MW_{SZN3} = 24,56 % (s = 5,20 %) für Szenario 3 und für die beiden Szenarien mit niedriger Belastung

geringere Werte mit MW_{SZN2} = 12,28 % (s = 3,57 %) für Szenario 2 und MW_{SZN4} = 8,14 % (s = 2,38 %) (vgl. Abbildung 52). Ein Kolmogorov-Smirnoff Test auf Normalverteilung, sowie ein Levene's Test auf Varianzhomogenität der Modelldaten ergaben keine signifikanten Ergebnisse, sodass Normalverteilung und Varianzhomogenität angenommen werden können. Eine Messwiederholungs-ANOVA wurde durchgeführt, bei der die einzelnen Szenarien als vier Stufen für die Beanspruchung angesehen wurden.

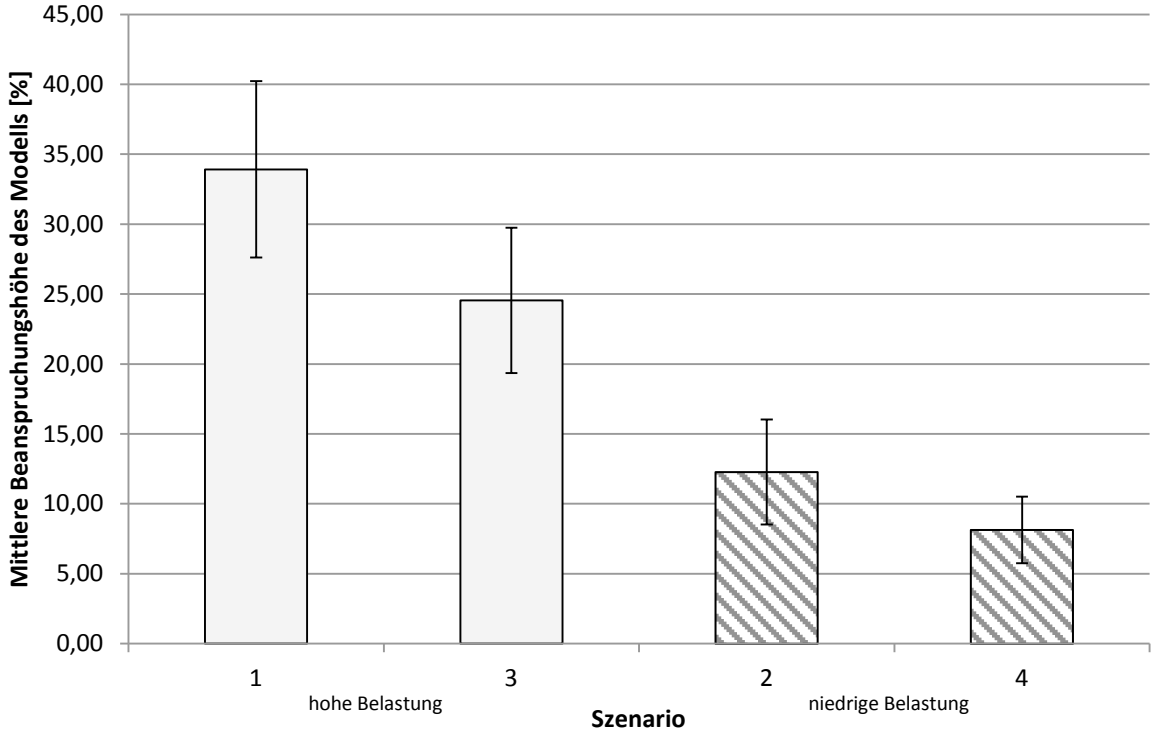

Abbildung 52: Mittleren Beanspruchungshöhe jedes Szenarios (in %) des Modells

Die Hauptanalyse ergab, dass sich die Beanspruchung hoch signifikant unterschiedet ($F(3, 69) = 153{,}79$; $p < 0{,}001$). Der Vergleich aller Szenarien untereinander mit Hilfe der geplanten Kontraste ergab hoch signifikante Unterschiede zwischen allen mittleren Beanspruchungswerten des Modells (p jeweils < 0,001).

Tabelle 15: Ergebnisse der Testung der Alternativhypothesen von Hypothese 7

Hypothese	Getätigte Annahme	Ergebnis
Hypothese 7a	Die mittlere Beanspruchungshöhe des Modells von SZN 1 ist höher als von SZN 2 (Mittelwertsdifferenz 21,639, p < 0,001, CI(95%) 17,463-25,815)	Bestätigt ✓
Hypothese 7b	Die mittlere Beanspruchungshöhe des Modells von SZN 1 ist höher als von SZN 4 (Mittelwertsdifferenz 25,778, p < 0,001, CI(95%) 21,960-29,596)	Bestätigt ✓

Hypothese 7c	Die mittlere Beanspruchungshöhe des Modells von SZN 3 ist höher als von SZN 2 (Mittelwertsdifferenz 12,278, p < 0,001, CI(95%) 8,191-16,364)	Bestätigt ✓
Hypothese 7d	Die mittlere Beanspruchungshöhe des Modells von SZN 3 ist höher als von SZN 4 (Mittelwertsdifferenz 16,417, p < 0,001, CI(95%) 12,810-20,024)	Bestätigt ✓

Es gelten damit die Alternativhypothesen 7a – d. Für die Überprüfung der Nullhypothesen wurde das α-Niveau analog zu oben wiederum mit α = 0,20 festgelegt. Es ergaben sich folgende Ergebnisse:

Tabelle 16: Ergebnisse der Testung der Nullhypothesen von Hypothese 7

Hypothese	Getätigte Annahme	Ergebnis
Hypothese 7e	Die mittlere Beanspruchungshöhe des Modells von SZN 1 und SZN 3 unterscheiden sich nicht (Mittelwertsdifferenz 9,361, p < 0,001, CI(95%) 4,716-14,006)	Nicht Bestätigt ✗
Hypothese 7f	Die mittlere Beanspruchungshöhe des Modells von SZN 2 und SZN 4 unterscheiden sich nicht (Mittelwertsdifferenz 4,139, p = 0,001, CI(95%) 1,591-6,687)	Nicht Bestätigt ✗

Die Nullhypothesen 7e und f, bei denen angenommen wurde, dass sich die Szenarien mit gleicher Belastung nicht signifikant voneinander unterscheiden, müssen abgelehnt werden, so dass auch hier Unterschiede in der mittleren Beanspruchung, die durch das Modell ermittelt wurde, bestehen.

6.2.8 Trendvergleich der mittleren Beanspruchungshöhe (Hypothese 8)

Hypothese (8): Es besteht eine starke Korrelation (ρ; τ > 0,6) zwischen der mittleren Beanspruchungshöhe der subjektiven (RSME und NASA-TLX) und der Modelldaten über die vier Szenarien.

Zum visuellen Vergleich der Trends der mittleren Beanspruchungshöhe über die Szenarien wurden die Prozentwerte der NASA-TLX-Sub-Skala geistige Anstrengung und der RSME-Bewertungen im Vergleich zu den Modelldaten dargestellt (vgl. Abbildung 53). Da bereits in Hypothese 5 (vgl. Kapitel 6.2.5) festgestellt wurde, dass sich die Trends über die Szenarien des NASA-TLX g.A. und des RSME unterschieden, wurde zusätzlich noch der Gesamtwert des NASA-TLX (Raw Task Load Index nach Byers et al., 1989; NASA-RTLX) bestimmt, um zu überprüfen, ob sich die Trends des RSME und des NASA-RTLX stärker ähneln. Hiermit sind Rückschlüsse auf die Diagnostizität des RSME möglich. Es ist zu erkennen, dass die absolute Beanspruchungshöhe

vom Modell zwar meist unterschätzt wird, sie sich aber in etwa auf der Höhe der mittleren Beanspruchungshöhe, die mit dem RSME erreicht wurde, befindet. Es ist ferner zu erkennen, dass der Trend, den die subjektive Bewertung mit der Sub-Skala „geistige Beanspruchung“ des NASA-TLX ergab, auch durch das Modell abgebildet werden kann (Szenario 1 > Szenario 3 > Szenario 2 > Szenario 4), was sich auch in einer hohen Korrelation zwischen den beiden Messskalen widerspiegelt ($\rho = 1{,}00$; $\tau = 1{,}00$).

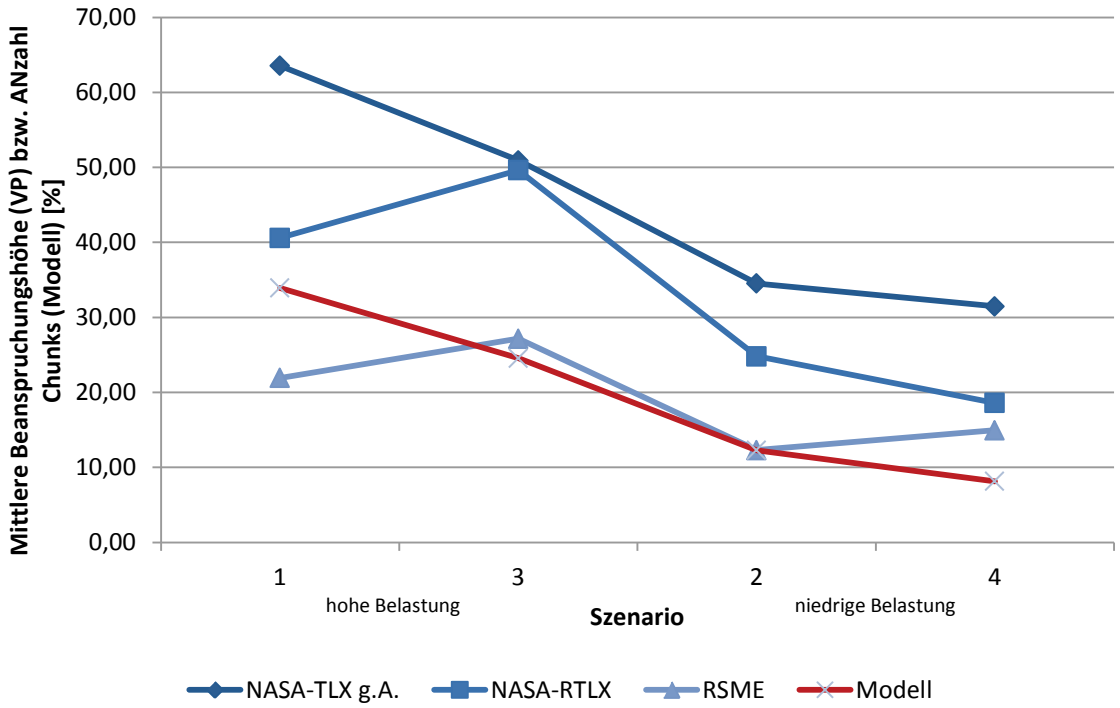

Abbildung 53: Mittlere Beanspruchungshöhe des Modells (Anzahl Chunks) verglichen mit der mittleren Beanspruchungshöhe des RSME, des NASA-TLX (g.A.), sowie des NASA-RTLX (jeweils in %)

In ähnlicher Weise wird auch der Trend der mittleren Fehlerzahlen durch das Modell abgebildet (vgl. Kapitel 6.2.3; Szenario 3 > Szenario 1 > Szenario 2 > Szenario 4; $\rho = 0{,}800$; $\tau = 0{,}667$). Die Trends des NASA-RTLX, sowie der RSME-Ratings bildet das Modell hingegen schlechter ab (RTLX: $\rho = 0{,}800$; $\tau = 0{,}667$; RSME: $\rho = 0{,}600$; $\tau = 0{,}333$). Es ist außerdem zu erkennen, dass der Trend des NASA-RTLX eher mit dem Trend der RSME-Bewertungen überreinstimmen (NASA-RTLX: Szenario 3 > Szenario 1 > Szenario 2 > Szenario 4; RSME: Szenario 3 > Szenario 1 > Szenario 4 > Szenario 2; $\rho = 0{,}800$; $\tau = 0{,}667$). Beide korrelieren allerdings schlechter mit der Sub-Skala „geistige Anstrengung“: (TLX g.A. – RTLX: $\rho = 0{,}800$, $\tau = 0{,}667$; TLX g.A. – RSME: $\rho = 0{,}600$, $\tau = 0{,}333$).

Die Hypothese 8, nämlich, dass eine starke Korrelation bezüglich der subjektiven und der Modelldaten über die vier Szenarien zu erwarten ist, gilt daher für den NASA-TLX g. A., die Fehlerzahl, sowie den NASA-RTLX. Mit beiden Koeffizienten ergaben sich starke Korrelationen (> 0,6). Lediglich für den RSME ergab sich mit Kendall’s τ nur eine schwache Korrelation, sodass

hier ein weniger starker Zusammenhang besteht. Tabelle 17 gibt einen Überblick über die einzelnen Zusammenhänge.

Tabelle 17: Übersicht der Korrelationskoeffizienten der einzelnen Messskalenvergleiche

Vergleich	**Spearman's ρ**	**Kendall's τ**
Modell – NASA-TLX g,A.	1,000	1,000
Modell – Fehlerzahl	0,800	0,667
Modell – NASA-RTLX	0,800	0,667
Modell – RSME	0,600	0,333
NASA-RTLX – RSME	0,800	0,667
NASA-TLX g.A. – NASA-RTLX	0,800	0,667
NASA-TLX g.A. – RSME	0,600	0,333

6.2.9 Überprüfung der Modellgüte mittels Einzelfallanalysen (Hypothese 9)

Hypothese (9): Es besteht ein positiver Zusammenhang zwischen den zeitlichen Verläufen der RSME- und Modelldaten.

Für die Beurteilung des Zusammenhanges zwischen den subjektiven Einschätzungen (mittels RSME) der einzelnen Versuchspersonen und den Modelldaten wurden zunächst visuelle Überprüfungen der zeitlichen Verläufe von Modelldaten und RSME-Daten sowie Streudiagramme herangezogen[68]. Die Quantifizierung wurde mit Hilfe der Korrelationskoeffizienten Kendall's τ und Spearman's ρ berechnet, wie in Kapitel 6.1.4 beschrieben. Hierzu sind Einzelfallanalysen der Versuchspersonenratings (alle Minuten-Werte eines Szenarios) durchgeführt und mit den Einzelfallsimulationen des Modells verglichen worden. Somit kann für jedes Versuchspersonenrating ein eigener visueller Vergleich, sowie ein eigener Korrelationskoeffizient erhalten werden. Dies dient einerseits der besseren Möglichkeit, bestimmte Besonderheiten der Ratings zu identifizieren, andererseits werden Hinweise erhalten, wie die Vorhersagegüte des Modells durch individuelle Unterschiede in den Versuchspersonendaten variiert. Weiterhin war die Einzelfallbetrachtung notwendig, da die Simulationssoftware es nicht erlaubte, exakt standardisierte Szenarien zu erstellen.

Stellvertretend soll hier das Streudiagramm und der Vergleich der zeitlichen Verläufe der Beanspruchung für die Versuchsperson 24 im Szenario 1 dargestellt werden. Die Abbildung 54 stellt das Streudiagramm der Versuchsperson 24 im Szenario 1 dar. Aus der Abbildung lässt sich ein hoher positiver korrelativer Zusammenhang erwarten. Abbildung 55 stellt die zeitlichen

[68] Ein zeitlicher Verlauf der mentalen Beanspruchung innerhalb der Szenarien wurde nur mit dem RSME erfasst, nicht mit dem NASA-TLX. Daher ist auch nur der Vergleich zwischen RSME und Modell möglich.

Verläufe der subjektiven Bewertungen sowie der Modelldaten ebenfalls für Versuchsperson 24 im Szenario 1 dar. Es ist zu erkennen, dass die Versuchsperson 24 gut durch das Modell abgebildet wird, sowohl der allgemeine Trend der Kurven, als auch Täler und Gipfel in den subjektiven Bewertungen.

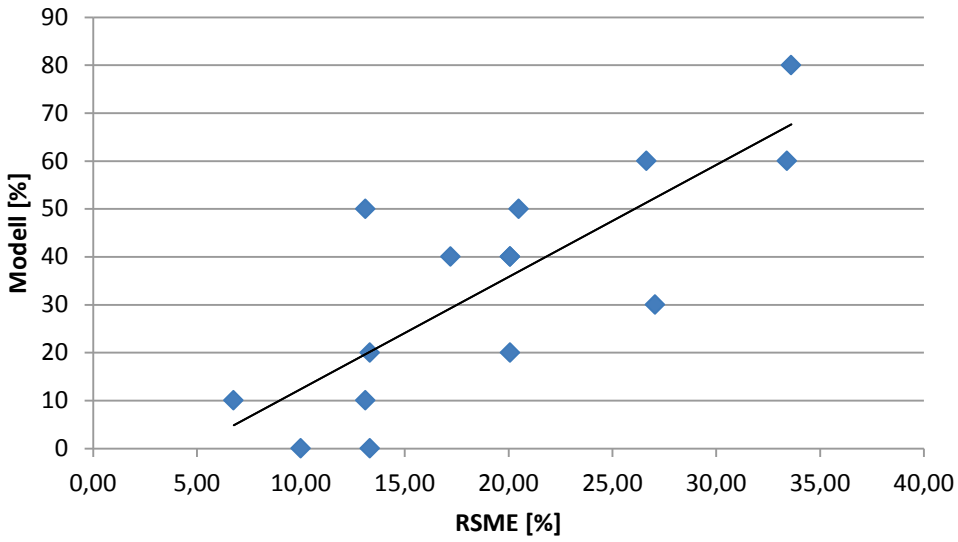

Abbildung 54: Streudiagramm Modell-RSME der Versuchsperson 24 im Szenario 1

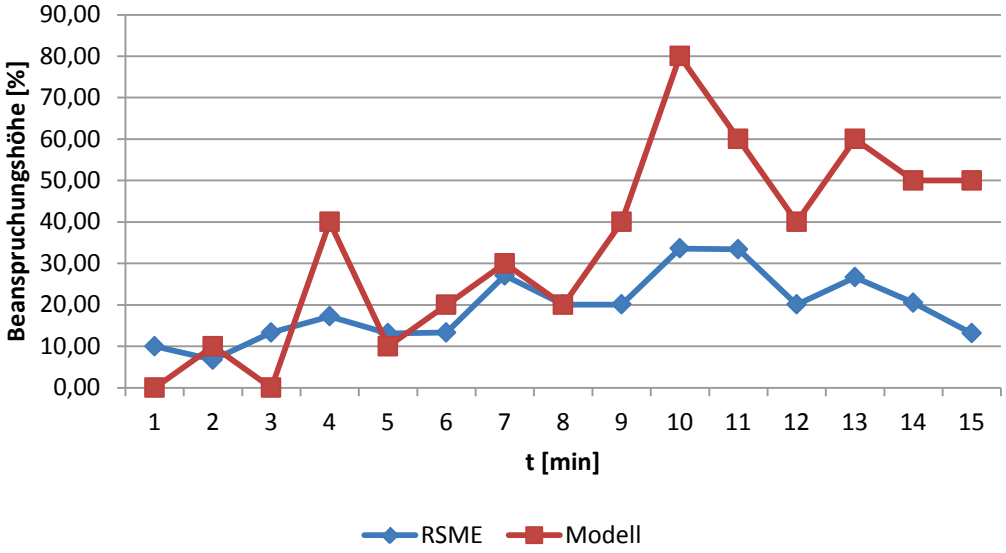

Abbildung 55: Zeitliche Verläufe der subjektiven Beanspruchungswerte (blau) sowie der Modellwerte (rot) der Versuchsperson 24 im Szenario 1

Es sieht allerdings so aus, dass das Modell auf Belastungsunterschiede stärker zu reagieren scheint, als die Versuchsperson. Somit ist die Kurve der Versuchsperson etwas flacher als die des Modells. Zu Zeiten hoher Belastung reagiert das Modell meist mit höherer Beanspruchung und höheren Ausschlägen, als es bei der Einschätzung der Versuchsperson der Fall ist. Es ist ebenfalls zu erkennen, dass zum Ende des Szenarios die Kurven etwas auseinanderdriften. Alle

anderen 95 Streudiagramme und Verläufe sind den im Anhang beigefügten Exceldateien zu entnehmen[69].

Tabelle 18: Zusammenfassung aller möglichen Korrelationen zwischen Modell und RSME pro Zeitreihe des Szenarios 1

VP	Kendall's τ	Spearman's ρ
4	-0,500	-0,679
5	0,322	0,445
6	0,565	0,734
11	0,493	0,634
13	0,618	0,732
14	0,617	0,792
15	0,418	0,530
16	0,060	0,103
17	0,279	0,368
18	0,325	0,496
19	0,269	0,366
20	0,411	0,514
21	0,663	0,810
22	0,382	0,516
24	0,640	0,751
Mittelwert	**0,392**	**0,518**

Es wurde anschließend für jede Versuchsperson individuell die zu den 15 Messzeitpunkten erhobenen subjektiven Bewertungen mit den Modellwerten der korrespondierenden 15 Modellzeitpunkte korreliert. Hierbei galt es allerdings eine Einschränkung zu machen: Einige der Fragebögen sind von den Versuchspersonen nicht ausgefüllt worden, weshalb diese aus der Analyse ausgeschlossen wurden[70]. Für jedes Szenario wurde daher beschrieben, wie viele der 24 möglichen Korrelationen tatsächlich in die Auswertung mit einbezogen werden konnten.

[69] Die in den Exceldateien enthaltenen Diagramme der zeitlichen Verläufe sind aufgrund der besseren Vergleichbarkeit und der Möglichkeit, die Zeitpunkte zu denen Fehler auftraten mit darzustellen, auf Basis z-standardisierter Daten erstellt worden, da wie bereits argumentiert die absolute Beanspruchungshöhe nicht so erheblich ist, wie der Beanspruchungsverlauf.

[70] Die Versuchspersonen waren angehalten, bei jedem auftretenden Piep-Ton den Fragebogen auszufüllen. Sollten sie auf Grund der Überlagerung des Piep-Tones mit Kommunikation oder durch zu hohe Beanspruchung einmal einen Fragebogen nicht angekreuzt haben, so kam es zu einer Auslassung der vornummerierten Bögen. Dies resultierte darin, dass eine Versuchsperson in Minute 5 Bogen 4 ausgefüllt hat und damit in Minute 6 Bogen 5 usw. Da während der Versuche keine Beobachtung der Versuchspersonen erfolgte, kann nicht rekonstruiert werden, an welcher Stelle der Zeitreihe sich der fehlende Wert tatsächlich befindet.

Repräsentativ für das erste Szenario sind die Korrelationskoeffizienten der Zeitreihen der einzelnen Versuchspersonen in Tabelle 18 dargestellt.

Die Tabellen der anderen Szenarien, sowie die ausführlichen Korrelationstabellen befinden sich im Anhang. Es kann festgehalten werden, dass für die subjektiven Bewertungen über alle Szenarien in ca. $\frac{2}{3}$ der Fälle Korrelationen berechenbar waren (64 von 96 insgesamt möglichen; 66,67 %). Von diesen berechenbaren Korrelationen zeigten 92,19 % (τ) bzw. 90,63 % (ρ) einen positiven Zusammenhang zwischen den Versuchspersonenratings und den Modelldaten. Lediglich 5 (7,81 %; τ) bzw. 6 (9,38 %; ρ) der berechenbaren Korrelationen waren negativ. Die Anzahl der möglichen, sowie der positiven Korrelationen sind in Tabelle 19 zusammengefasst.

Tabelle 19: Zusammenfassung der möglichen und positiven Korrelationen pro Szenario und insgesamt, sowie deren prozentualer Anteil

Szenario	Anzahl möglicher Korrelationen	Anzahl positiver Korrelationen (τ)	Anzahl positiver Korrelationen (ρ)
1	15	14	14
2	18	16	15
3	13	12	12
4	18	17	17
Gesamt	**64**	**59**	**58**
Prozentualer Anteil (gesamt)	**66,67**	**92,19**	**90,63**

Die Berechnung der mittleren Korrelationen für jedes Szenario erfolgte mit Hilfe einer Fisher Z-Transformation (vgl. Bortz & Schuster, 2010), so dass sich jeweils eine schwach bis mittlere Korrelation zwischen dem RSME und den Modelldaten ergab (vgl. Tabelle 20).

Tabelle 20: Mittlere Korrelationen aller Zeitreihen pro Szenario, sowie über alle Szenarien

Szenario	Kendall's τ	Spearman's ρ
1	0,392	0,518
2	0,243	0,305
3	0,323	0,410
4	0,382	0,461
Mittelwert	**0,336**	**0,427**

Im Mittel über alle Szenarien besteht nach Spearman's ρ ein mittlerer Zusammenhang (ρ = 0,427), nach Kendall's τ ein schwacher Zusammenhang (τ = 0,392). Damit gilt die Hypothese 9. Aufgrund der in Hypothese 2 festgestellten geringen Sensitivität der Herzrate kann für die

Korrelation zwischen den Modelldaten und den EKG-Daten davon ausgegangen werden, dass kein Zusammenhang besteht, da die Hypothese 10 nicht mehr valide ist.

6.2.10 Explorative Analyse: Absolute Abweichung der Modelldaten

Exploratives Vorgehen: Berechnung der absoluten (AD) und mittleren absoluten Abweichung (MAD) der mittleren Beanspruchungshöhe des Modells von den Realdaten (RSME und NASA-TLX) innerhalb jedes Szenarios.

Um die Abweichung der Modelldaten von den experimentellen Daten bestimmen zu können, soll im Folgenden zunächst für jedes Szenario die absolute Abweichung sowie die mittlere absolute Abweichung (MAD) berechnet werden. Dies soll für beide subjektive Maße erfolgen.

Wie bereits in der Abbildung 53 (vgl. Kapitel 6.2.8) zu erkennen war, liegen die Modellwerte eher auf Höhe des RSME. Die absolute Abweichung ergibt sich aus der Differenz der Mittelwerte (bzw. Mediane) zwischen Modell- und subjektiven Daten. Die absolute Abweichung des Modells vom RSME beträgt für Szenario 1 AD_{SZN1} = 11,99 %, für Szenario 2 AD_{SZN2} = -0,02 %, für Szenario 3 AD_{SZN3} = -2,60 % und für Szenario 4 AD_{SZN4} = 6,82 %. Die mittlere absolute Abweichung ergibt sich nach der Formel (vgl. Schunn & Wallach, 2005, S. 24):

$$MAD = \frac{\sum_{i=1}^{k} |m_i - d_i|}{k},$$

wobei m_i den Modellmittelwert jedes Szenarios i, d_i den Datenmittelwert jedes Szenarios und k die Anzahl der verglichenen Szenarien wiedergibt. Die mittlere absolute Abweichung der Modellwerte vom RSME ergibt sich damit zu MAD = 5,36 %. Die absolute Abweichung von NASA-TLX g. A. und Modell beträgt für Szenario 1 AD_{SZN1} = -29,62 %, für Szenario 2 AD_{SZN2} = -22,24 %, für Szenario 3 AD_{SZN3} = -26,37 % und für Szenario 4 AD_{SZN4} = -23,34 %. Die mittlere absolute Abweichung ergibt sich zu MAD = -25,39 %. Die mittlere Beanspruchungshöhe, die mit dem NASA-TLX g. A. erreicht wurde, wird damit vom Modell weit unterschätzt, wobei die Werte des RSME recht gut abgebildet werden können.

Tabelle 21: Absolute und mittlere Absolute Abweichung der Modelldaten von den subjektiven Daten (RSME und NASA-TLX)

Szenario	**Absolute Abweichung RSME [%]**	**Absolute Abweichung NASA-TLX [%]**
1	11,99	-29,62
2	-0,02	-22,24
3	-2,60	-26,37
4	6,82	-23,34
Mittlere absolute Abweichung	5,36	25,39

6.2.11 *Explorative Analyse: Fehler und Beanspruchungspeaks*

Exploratives Vorgehen: Berechnung der relativen Häufigkeit, mit der Fehler von Versuchspersonen zu Zeitpunkten auftreten, in denen das Modell eine hohe Beanspruchung vorhersagt.

Die explorative Analyse geht davon aus, dass eine erhöhte Fehlerwahrscheinlichkeit besteht, wenn die Beanspruchung der Versuchspersonen hoch ist. Hierbei galt es festzustellen, ob aufgetretene Fehler häufig innerhalb solcher Beanspruchungsgipfel liegen. Es ist dabei möglich eine Vierfeldertafel laut Signalentdeckungstheorie (Green & Sweets, 1966) aufzustellen (vgl. Tabelle 22). Dabei ist allerdings zu beachten, dass ein Beanspruchungsgipfel nicht zwangsläufig mit einem Fehler einhergehen muss, sondern sich lediglich seine Auftretenswahrscheinlichkeit erhöht. Es stellt sich damit als schwierig heraus, sowohl falsche Alarme (false Alarm) als auch korrekte Zurückweisungen (correct rejection) zu beurteilen. Informativ ist daher nur der Anteil an Treffern (Hit) und Auslassungen (Miss). Außerdem ist zu beachten, dass nur in 20 % der Zeit Fehler gemacht wurden. 24 Versuchspersonen arbeiteten jeweils vier Szenarien mit je 15 Minuten Dauer ab, was 1440 Minuten entspricht. Insgesamt wurden dabei von allen Versuchspersonen in allen Szenarien 300 Fehler gemacht.

Tabelle 22: Vierfeldertafel der Signalentdeckungstheorie bezogen auf die Fehlerzahl in Versuch und Modell

VP\Modell	**Fehler**	**Kein Fehler**
Peak	Treffer	falscher Alarm
Kein Peak	Auslassung	korrekte Zurückweisung

Die Wahrscheinlichkeit eines Treffers (pHit; probability of hit) ergibt sich dabei aus der Anzahl an Treffern, geteilt durch die Anzahl an Treffern plus der Anzahl an Auslassungen ($pHit = \frac{His}{Hits+Miss}$). Wobei sich die Wahrscheinlichkeit einer Auslassung durch 1-Treffer ergibt. Es wurde zur Auswertung folgende Kriterien festgelegt und untersucht: Anhand der aufgetretenen Fehler wurde untersucht, ob dieser innerhalb eines lokalen Maximums, bzw. innerhalb eines lokalen Maximums mit einer Toleranz von ± 1 Minute der Modelldaten liegt. Zusätzlich zu diesem Kriterium, wurde definiert, dass ein Modellwert über fünf (> 5 Chunks im Arbeitsgedächtnis) eine erhebliche Beanspruchung bedeutet, wodurch sich hier auch die Wahrscheinlichkeit des Auftretens von Fehlern im Experiment erhöhen sollte. Es wurde daher festgelegt, dass immer wenn ein Fehler der Versuchsperson auftritt und entweder eines der beiden oder beide Kriterien zugleich erfüllt waren, dies als Treffer gezählt wurde. Alles andere galt als Auslassung.

Es konnte dabei festgestellt werden, dass im Szenario 1 57 % der Fehler innerhalb von Beanspruchungsgipfeln liegen. Das bedeutet im Umkehrschluss, dass 43 % der Fehler nicht innerhalb solcher Gipfel liegen. Im Szenario 2 konnten sogar 67 % der Fehler innerhalb von

Beanspruchungsgipfeln festgestellt werden (33 % nicht). Die Szenarien 3 und 4 ergaben jeweils 62 % (38 % nicht) bzw. 58 % (42 % nicht).

Dies bedeutet, dass im Mittel 61 % der Fehler innerhalb von Beanspruchungsgipfeln lagen, die das Modell vorhersagte, 39 % der Fehler allerdings nicht. Es konnten also 183 der insgesamt 300 Fehler vorhergesagt werden. Die Ergebnisse sind in Tabelle 23 zusammengefasst.

Tabelle 23: Zusammenfassung der Ergebnisse: Fehler innerhalb von Beanspruchungsgipfeln

Szenario	In Beanspruchungs-gipfeln	Nicht in Bean-spruchungsgipfeln
1	57 %	43 %
2	67 %	33 %
3	62 %	38 %
4	58 %	42 %
Mittel	61 %	39 %

6.3 Diskussion der Ergebnisse

6.3.1 Zusammenhang d2-R Ergebnis und Fehlerzahl bzw. Höhe der Beanspruchungsratings (Hypothese 1)

Allgemein wurde davon ausgegangen, dass der Effizienz-Gründlichkeits-Tradeoff, mit dem Versuchspersonen kognitive Aufgaben bearbeiten, einen Einfluss darauf hat, wie viele Fehler in den Szenarien gemacht werden bzw. wie hoch die Beanspruchung bei der Aufgabenbearbeitung eingeschätzt wird. Es wurde angenommen, dass mit steigendem d2-R-Wert der Versuchspersonen (besserem Abschneiden im Test) die Fehlerzahl abnimmt. Gleichzeitig wurde vermutet, dass der Effizient-Gründlichkeits-Tradeoff besser gelingt, wodurch auch die Beanspruchung mit steigendem d2-R-Wert als geringer eingeschätzt werden würde.

Hinsichtlich der Fehlerzahl ergibt sich zumindest deskriptiv das argumentierte Bild. Es konnte gezeigt werden, dass mit steigendem d2-R-Wert die Fehlerzahl abnimmt. Zwar ist dieses Ergebnis nicht statistisch signifikant, es weist aber dennoch darauf hin, dass zumindest ein kleiner Zusammenhang besteht. Wie in der Hypothese angenommen, gelingt es den Versuchspersonen, die im d2-R Test gut abschneiden anscheinend auch in der Fluglotsenaufgabe die relevanten Zielreize schnell zu selektieren und ebenso schnell zu guten Entscheidungen zu kommen. Versuchspersonen mit schlechterem d2-R Ergebnis hingegen brauchen länger zur Selektion der relevanten Reize, wodurch öfter kritische Situationen entstehen, die zu Fehlern führen.

Beim Vergleich der drei Gruppen (hoher, durchschnittlicher und geringer d2-R-Wert) hinsichtlich ihrer abgegebenen Beanspruchungsbewertungen ist mit sehr viel höherer Sicherheit

davon auszugehen, dass kein statistisch bedeutsamer Unterschied der Beanspruchungsbewertungen der drei Gruppen vorliegt. Interessant ist hierbei allerdings das deskriptive Ergebnis, dass die Gruppe mit durchschnittlichem d2-R-Wert die Beanspruchung im Mittel als geringer einschätzte als die beiden anderen Gruppen (durchschnittlich < niedrig < hoch). Die Versuchspersonen mit durchschnittlichem d2-R-Wert fühlten sich in der Fluglotsenaufgabe also am wenigsten beansprucht. Dieses Bild ergab sich für beide subjektiven Maße. Die Aufgaben in der simulierten Fluglotsenaufgabe sind dabei mit den Aufgaben des Konzentrationstests zu vergleichen: Es muss auf (relativ) einfache Reize reagiert werden, die eindeutig wahrnehmbar sind, wobei einfache Regeln anzuwenden und Teilhandlungen so schnell wie möglich zu koordinieren sind. Hierbei kann vermutet werden, dass die Versuchspersonen mit hohem d2-R-Wert eine andere Strategie verfolgten als die anderen beiden Gruppen. So könnten die Versuchspersonen mit geringem, als auch mit durchschnittlichem Abschneiden im d2-R-Test den Flugverkehr nach den vorgegebenen Regeln abgearbeitet haben, was mit relativ geringem kognitivem Aufwand zu erfüllen ist. Die Gruppe mit durchschnittlichem d2-R-Wert empfand dabei das Reagieren auf die Reize und die Anwendung der Regeln als weniger beanspruchend als die Gruppe mit geringem d2-R-Wert, aufgrund der Tatsache, dass relevante Zielreize unter Zeitdruck besser extrahiert werden konnten.

Für die Versuchspersonen mit hohem d2-R-Wert kann vermutet werden, dass ihnen die reine Selektion der Zielreize und Regelanwendung noch leichter fällt und sie damit über mehr freie kognitive Ressourcen verfügen. Sie sind daher in der Lage, diese freien Ressourcen zu verwenden, um komplexere Problemlösestrategien anzuwenden und beispielsweise Verkehrssituationen in die Zukunft zu projizieren, um bessere Entscheidungen treffen zu können. Da diese Strategie allerdings einen größeren Ressourcenbedarf erfordert, ist die empfundene mentale Beanspruchung dieser Versuchspersonen höher, als die Beanspruchung der Versuchspersonen, die lediglich durch die Anwendung einfacher Regeln auf Reize reagieren. Hiermit kann auch wiederum die geringe Fehlerzahl der Versuchspersonen mit hohem d2-R-Wert erklärt werden. Da aufgrund der angewendeten Strategie kritische Situationen vorhergesehen werden können, werden bessere Entscheidungen getroffen, wodurch weniger Fehler entstehen, als bei reiner Regelanwendung.

Das Abschneiden im d2-R-Test wurde aufgrund der nicht signifikanten Ergebnisse nicht als Kontrollvariable mit in die folgenden Analysen einbezogen.

6.3.2 Bewertung der Sensitivität der Herzrate (Hypothese 2)

Bei den Untersuchungen zur Hypothese 2 konnte festgestellt werden, dass die Herzrate nicht sensitiv genug ist, um zwischen zwei Experimentalbedingungen zu unterscheiden. Allgemein

wurde angenommen, dass die Unterschiede in der Belastung der Szenarien mit hohem und niedrigem Verkehrsaufkommen groß genug seien, dass die Herzrate in der Lage ist, diese Beanspruchungsunterschiede zu differenzieren. Wie sich zeigte ist dies nicht der Fall, wodurch die Befunde, welche Nickel et al. (2002) berichten, bestätigt werden konnten. Auch hier wurde gezeigt, dass die Herzrate nicht zwischen verschiedenen Belastungsbedingungen sensitiv unterscheiden kann, sondern nur zwischen Belastungs- und Baselinebedingungen. Ein Nutzen der Herzrate für weitere Analysen ist damit nicht gegeben, da kein Informationsgewinn zu erwarten ist.

Die Ursachen für die geringe Sensitivität könnten darin liegen, dass die kardiovaskulären Reaktionen der Versuchspersonen auf die experimentelle Manipulation, womöglich sogar auf mentale Beanspruchung, sehr unterschiedlich sind. Gemeinhin wird angenommen, dass mit steigender mentaler Beanspruchung eine Steigerung der Herzschläge pro Minute (bpm) und damit ein Absinken der Abstände der Herzschläge (IBI) einhergeht, wie es auch in Hypothese 2 formuliert worden ist (Manzey, 1998; Roscoe, 1992, 1993; Vogt et al., 2006). Befunde sprechen allerdings auch dafür, dass die längere Exposition in einer Aufgabe zu einer Art Adaptionsreaktion führt, die ein Absinken der Herzschläge pro Minute und damit einen Anstieg der Inter-Beat-Abstände zur Folge hat (Hörhold & Walschburger, 1998; Hörhold, 1994). Durch die 15-minütige Dauer eines jeden Szenarios könnte eine solche Anpassungsreaktion aufgetreten sein, sodass das initiale Absinken der Inter-Beat-Abstände durch den Anstieg im Verlauf des Szenarios neutralisiert worden ist und somit im Mittel keine Veränderungen der Inter-Beat-Abstände beobachtet werden können.

Ebenso kann auch die sogenannte Laceys Intake-Rejection-Hypothese (Lacey & Lacey, 1978; Lacey, 1967) eine Erklärung für diese Ergebnisse sein. Diese besagt, dass die Aufnahme von Informationen aus der Umwelt (visuelle Wahrnehmung, Hören) mit einer Abnahme, die Verarbeitung und Abgabe von Informationen (Kopfrechnen, Gedächtnisabruf, Problemlösen) mit einem Anstieg der Herzrate verbunden ist (vgl. Kramar, 1993; Ribback, 2003). Beide Effekte könnten sich somit gegenseitig aufgehoben haben.

Es kann weiterhin davon ausgegangen werden, dass die Beanspruchung von Lotsen an ihrem realen Arbeitsplatz sehr viel höher ist, als es in einer simulierten Umgebung der Fall ist (De Waard, 1996). Ebenso kann es möglich sein, dass die Herzrate durch andere Faktoren beeinflusst wurde. Neben physischer Anstrengung, die im vorliegenden Versuch eher gering war, können hier vor allem auch emotionale Faktoren eine Rolle gespielt haben. Große Verantwortung oder die Angst Fehler zu machen, wie sie bei Jorna (1992) berichtet werden, können auch im vorliegenden Versuch eine Rolle gespielt haben. Es kann davon ausgegangen werden, dass beide emotionale Faktoren in allen Szenarien gleichermaßen vorlagen und somit

die Herzrate der Versuchspersonen beeinflussten, wobei diese den Effekt der mentalen Beanspruchung zu einem gewissen Grad überlagert haben könnten.

6.3.3 Bewertung der mittleren mentalen Beanspruchung erfasst durch die RSME-Skala (Hypothese 4)

Die Hypothesentestung zur Feststellung der experimentellen Manipulation mit dem RSME kann als erfolgreich angesehen werden. So gelten alle Alternativhypothesen. Eine der beiden Nullhypothesen ergab signifikante Ergebnisse, so dass festgestellt wurde, dass sich Szenario 1 und 3 in ihrer Beanspruchung unterscheiden. Auch zwischen den Szenarien 2 und 4 sind deskriptiv Unterschiede zu erkennen. Die Ursache dafür, dass geringfügige Unterschiede der Bewertung mentaler Beanspruchung in Szenarien mit gleicher Belastung (gleiche Anzahl Flugzeuge insgesamt) auftreten, könnten in der Unterschiedlichen Verteilung der Flugzeuge über die Zeit liegen. So könnten hierdurch unterschiedlich schwere Verkehrssituationen zustande gekommen sein, die von den Versuchspersonen mit veränderter Beanspruchungshöhe wahrgenommen wurden. Während also in einem Szenario die Flugzeuge so verteilt waren, dass sich beispielsweise meist nur zwei Flugzeuge, in größerer Entfernung voneinander gleichzeitig auf dem Flughafen befanden, könnten im anderen Szenario häufiger Situationen aufgetreten sein, in dem sich zwei Flugzeuge in einer potentiellen Konfliktsituation befanden. Es liegt hierbei eine Konfundierung der Variablen „Anzahl Flugzeuge“ mit Variablen wie „Anzahl Anfragen/Freigaben“, „Anzahl zu treffender Entscheidungen“ oder „Schweregrad der Entscheidungen“ vor, welche allerdings nicht ohne weiteres aufzulösen ist. Durch eine Erhöhung des Verkehrsaufkommens oder einer veränderten Verteilung der Flugzeuge über die Zeit, kommt es zwangsläufig immer dazu, dass mehr Kommunikation stattfindet, mehr Entscheidungen zu treffen sind und mehr potentielle Konfliktsituationen auftreten können. Die Verteilung der Flugzeuge über die Zeit war allerdings einerseits durch die Beschaffenheit der Simulation, andererseits aber auch durch den Einfluss der Versuchspersonen auf den Ablauf des Szenarios äußerst schwer zu kontrollieren.

6.3.4 Bewertung der Mittleren mentalen Beanspruchung erfasst durch den NASA-TLX (Hypothese 5)

Die Hypothesentestung zur Feststellung der experimentellen Manipulation mit dem NASA-TLX g. A. kann ebenfalls als erfolgreich angesehen werden. Hierbei gelten alle Alternativ- sowie Nullhypothesen. Es konnte festgestellt werden, dass hypothesenkonform, diejenigen Szenarien mit hoher Belastung auch zu höheren Beanspruchungsratings führten als Szenarien mit niedriger Belastung. Szenarien mit gleicher Belastung unterscheiden sich deskriptiv, jedoch nicht statistisch signifikant. Auch hierbei kann die unterschiedliche Beanspruchungshöhe der

Szenarien mit gleicher Belastung durch die veränderte Verteilung der Flugzeuge erklärt werden. Die Argumentation ergibt sich analog zu oben.

Interessant ist das Ergebnis, dass sich mit dem NASA-TLX g. A. eine etwas andere Einschätzung der mittleren mentalen Beanspruchung über die vier Szenarien ergibt, als mit dem RSME. So ist beispielsweise mit dem NASA-TLX das Szenario 1 als am beanspruchendsten eingeschätzt worden, mit dem RSME jedoch das Szenario 3. Dieses Ergebnis wird in der Diskussion zu Hypothese 8 (Kapitel 6.3.7) ausführlich diskutiert.

Ebenfalls fällt auf, dass die Beanspruchung sehr viel höher (meist doppelt so hoch) eingeschätzt wurde, als mit dem RSME. Dies kann einerseits durch die Instruktion und andererseits durch den unterschiedlichen Erhebungszeitpunkt der beiden Skalen erklärt werden. Die Versuchspersonen waren instruiert mit der RSME-Skala ihre aktuelle Beanspruchung einzuschätzen, die sie zu dem bestimmten Zeitpunkt innerhalb des Szenarios empfanden, an denen der Piep-Ton ertönte. Mit dem NASA-TLX sollten sie allerdings einen Gesamteindruck ihrer Beanspruchung über das gesamte Szenario widergeben. Hierbei ist es wahrscheinlich, dass die Versuchspersonen das gesamte Szenario als hoch beanspruchend bewerteten, jede einzelne Bewertung innerhalb des Szenarios allerdings eher im mittleren Bereich ausfiel. Ebenso können Verzerrungen dadurch aufgetreten sein, wie häufig bzw. in welcher Reihenfolge die Fragebögen angewendet wurden. Es wäre interessant gewesen auch den Gesamteindruck über das Szenario nochmals mit der RSME-Skala bewerten zu lassen, um einen Vergleich zu erhalten und so Rückschlüsse daraus ziehen zu können, ob die entstandenen Unterschiede durch unterschiedliche Messzeitpunkte und die unterschiedliche Instruktion zu erklären sind oder ob durch die Art der angewendeten Skala unterschiedliche Bewertungen zustande gekommen sind.

Letzteres kann ebenso die unterschiedlichen Ergebnisse erklären: Beim NASA-TLX wird eine 100 mm lange Skala ohne Ziffernangeben mit nur zwei verbalen Ankern (gering und hoch) an den jeweiligen Skalenenden verwendet. Beim RSME hingegen wird eine 244 mm lange Skala mit einer Ziffernkodierung in Zehnerschritten von 0 bis 150 verwendet. Zusätzlich ist die Skala mit zehn verbalen Ankern versehen und farblich kodiert (zunehmender Rotton je höher die Zahlenwerte werden). Die Präsentation von Zahlenwerten und verbalen Ankern könnte dazu geführt haben, dass die Versuchspersonen besser[71] in der Lage waren, die von ihnen empfundene Beanspruchungshöhe einzuordnen. Die Länge der RSME-Skala könnte hierbei dazu geführt haben, dass die Versuchspersonen ihre Beanspruchung genauer differenzieren konnten.

[71] Ob die Versuchspersonen wirklich besser in der Lage waren ihre Beanspruchung mit dem RSME einzuschätzen wird lediglich vermutet. Sicher ist allerdings, dass die genannten Unterschiede in der Beschaffenheit der Skalen auch Unterschiede in der Bewertungshöhe mit sich bringen, sodass die Versuchspersonen ihre Beanspruchung auf andere Weise einschätzen.

Auch die farbliche Gestaltung könnte eine zusätzliche Hilfe gewesen sein, die Beanspruchung besser einzuordnen.

Die Unterschiede können aber ebenso auf die Berechnungs- und Erfassungsmethode der mittleren Beanspruchungshöhe des jeweiligen Maßes zurückgeführt werden. Während der NASA-TLX nur ein Mal pro Szenario (nämlich am Ende) erfasst wurde, ermitteln sich die mittleren Beanspruchungswerte der RSME-Skala aus dem Median aller Versuchspersonen zu allen Messzeitpunkten, wobei es sich um eine „echte“ mittlere Beanspruchung handelt. Beim RSME fließen einerseits mehr Messwerte in die Berechnung ein (15 pro Versuchsperson), andererseits treten sowohl hohe als auch geringe Werte auf, die im Mittel eine mäßige Beanspruchung für das Gesamtszenario ergeben. Beim NASA-TLX wird die mittlere Beanspruchung nur durch den einen pro Versuchsperson ermittelten Messwert widergespiegelt, der am Ende des Szenarios durch die Versuchspersonen abgegeben wurde. Es kann weiterhin vermutet werden, dass in die einzelne Bewertung jedes Szenarios mittels NASA-TLX Zeiten hoher Beanspruchung innerhalb eines Szenarios stärker in die Gesamtbewertung der Versuchspersonen einfließen, als Zeiten geringer Beanspruchung. Hierbei sind beanspruchende Situationen, die innerhalb des Szenarios auftraten am Ende stärker repräsentiert, als wenig beanspruchende.

6.3.5 Bewertung der mittleren Fehlerzahl (Hypothese 6)

Auch hinsichtlich der Fehlerzahl kann die Hypothesentestung zur Feststellung der experimentellen Manipulation als erfolgreich betrachtet werden. So gelten alle Alternativhypothesen, sowie eine Nullhypothese. Für die Hypothese 6f muss davon ausgegangen werden, dass sich die mittlere Fehlerzahl der Szenarien 2 und 4 unterscheiden.

Grund hierfür, dass geringfügige Unterschiede der Fehlerzahl in Szenarien mit gleicher Belastung (gleiche Anzahl Flugzeuge insgesamt), ebenso wie bei den subjektiven Bewertungen, entstehen, kann auch hier in der unterschiedlichen Verteilung über die Zeit liegen. So könnten durch die unterschiedliche Verteilung der Flugzeuge unterschiedlich schwere Verkehrssituationen zustande gekommen sein, die von einigen Versuchspersonen nicht adäquat gelöst werden konnten, sodass es zu Fehlern kam. Die Argumentation ergibt sich analog zum Kapitel 6.3.3.

Interessant ist hierbei wiederum das Ergebnis, dass sich mit den mittleren Fehlerzahlen eine etwas andere Einschätzung der mittleren mentalen Beanspruchung ergibt, als mit dem RSME und dem NASA-TLX. So sind im Szenario 3 die meisten Fehler gemacht worden, im Szenario 4 die wenigsten (Szenario 3 > Szenario 1 > Szenario 2 > Szenario 4). Die Bewertungen sowohl mit dem NASA-TLX g. A. als auch mit dem RSME zeigen ein etwas anderes Bild (Szenario 1 > Szenario 3 > Szenario 2 > Szenario 4 mit NASA-TLX g. A.; Szenario 3 > Szenario 1 > Szenario 4 >

Szenario 2 mit RSME). Hierbei kann vermutet werden, dass durch die Fehlerzahl andere Faktoren mentaler Beanspruchung abgebildet werden, als sie mit den subjektiven Methoden (RSMA und NASA-TLX g. A.) erfasst worden sind. Tatsächlich ergab sich durch die explorative Auswertung der NASA-TLX-Skala „Frustration“ das gleiche Bild, wie es sich in den Fehlerzahlen widerspiegelt (Szenario 3 > Szenario 1 > Szenario 2 > Szenario 4). Es kann daher vermutet werden, dass die Fehlerzahl wiederum einen anderen Aspekt der Beanspruchung und nicht explizit mentale Beanspruchung widerspiegelt. Hierbei könnte durch die Fehlerzahlen eher eine Art Überforderung abgebildet werden, was eventuell sogar auf die maximale Arbeitsgedächtniskapazität der Versuchspersonen hinweisen könnte. Dies ist allerdings etwas schwierig zu beurteilen, da auch Unterforderung zu Fehlern führen kann. So konnte während der Versuchsdurchführung beobachtet werden, dass einzelne Versuchspersonen innerhalb eines gering belastenden Szenarienabschnittes Fehler machten. Diese könnten durch Unachtsamkeit entstanden sein, die mit der geringen Belastung einherging. Es konnte weiterhin bei einer Versuchsperson beobachtet werden, dass sie sich, unabhängig von der Instruktion, in Phasen geringer Belastung eine Nebenaufgabe suchte, indem sie den Versuchsleitern mitteilte, welche Flugzeuge ihre auf dem Flugstreifen ausgewiesene Ankunfts- oder Abflugzeit bereits überschritten hatten und damit verspätet eintreffen würden. Die Versuchsperson könnte hiermit eine Strategie angewendet haben, um einer eventuellen Unterforderung entgegenzusteuern. Dieses Verhalten kann aber ebenso zu Fehlern geführt haben, wenn die Aufmerksamkeit der Versuchsperson auf der irrelevanten Nebenaufgabe liegt und für die Versuchsperson unerwartet wieder auf die Hauptaufgabe gerichtet werden muss. Hierdurch kann die aktuelle Verkehrssituation nicht mehr genügend repräsentiert sein, sodass es zu Fehlern kommt. So ist zu erklären, dass das Szenario 2 mit dem NASA-TLX g. A. als am geringsten mental beanspruchend bewertet worden ist, aber mehr Fehler auftraten als in Szenario 4.

6.3.6 Mittlere Beanspruchungshöhe der Modelldaten (Hypothese 7)

Eine Signifikanzprüfung der Modelldaten brachte signifikante Unterschiede für alle Szenarien hervor, so dass die Unter-Hypothesen 7a-d gelten, die Hypothesen 7e und f abgelehnt werden müssen. In den Modelldaten unterschieden sich demnach auch die Szenarien mit gleicher Belastung signifikant in ihrer Beanspruchung. Diese Unterschiede können allerdings einerseits mit Unterschieden in der Verteilung der Flugzeuge über die Zeit innerhalb der Szenarien, andererseits mit der geringeren Varianz der Modelldaten erklärt werden. Bei geringer Varianz kann die einfaktorielle Varianzanalyse mit Messwiederholung bereits bei kleineren Mittelwertsunterschieden signifikante Ergebnisse hervorbringen. Die geringe Varianz der Modelldaten liegt darin begründet, dass keine interindividuellen Unterschiede, wie unterschiedliches Wissen, Strategien oder Arbeitsgedächtniskapazität, abgebildet werden, sondern ein- und dasselbe Modell aller Szenarien abarbeitet.

Deskriptiv waren diese Unterschiede zwischen den Szenarien auch in den Fragebogendaten zu erkennen, wobei sie mit der vorliegenden Stichprobengröße meist nicht die Signifikanzgrenze erreichten. Dieselben Unterschiede sind auch im Modell erkennbar, sodass geschlussfolgert werden kann, dass auch das Modell nicht nur die Unterschiede in der Beanspruchung aufdeckt, die durch die Anzahl der Flugzeuge entstehen, sondern auch die Unterschiede, die durch die veränderte Verteilung der Flugzeuge über die Zeit zustande kommen.

6.3.7 Trendvergleich der durch subjektive und leistungsbasierte Methoden erfassten Beanspruchungshöhe mit dem Modell (Hypothese 8)

Für die Modelldaten konnte ein vergleichbarer Trend festgestellt werden, wie er durch die Auswertung der Sub-Skala „geistige Anstrengung" des NASA-TLX widergespiegelt wird, was auch die hohe Korrelation (ρ und τ = 1,00) bestätigen konnte. Außerdem ist ein ähnlicher Trend wie bei den mittleren Fehlerzahlen ersichtlich. Es kann geschlussfolgert werden, dass das Modell in der Lage ist, mentale Beanspruchung diagnostisch abzubilden.

Hinweise hierauf konnten durch den Vergleich mit den Daten des NASA-RTLX gefunden werden: Hierbei wird der Trend des NASA-RTLX durch das Modell weniger gut wiedergegeben (geringere Korrelation) als der des NASA-TLX g. A. Da mit dem NASA-RTLX auch andere Beanspruchungsarten erfasst werden (z.B. werden körperliche Anforderungen durch ein Item erfasst), ist der Zusammenhang sowohl mit dem NASA-TLX g. A. als auch mit den Modelldaten geringer. Es ist weiterhin zu erkennen, dass der Trend der RSME-Bewertungen eher mit dem des NASA-RTLX übereinstimmt, aber weniger mit dem NASA-TLX g. A. Dies könnte ein Hinweis darauf sein, dass bei der Bewertung mittels RSME-Skala nicht nur die mentale Beanspruchung durch die Versuchspersonen bewertet wird, sondern hierbei auch andere Faktoren, die die wahrgenommene mentale Beanspruchung beeinflussen (z.B. Zeitdruck) mit in die Bewertung einfließen, da die Versuchspersonen nicht in der Lage sind dies in der Belastungssituation zu differenzieren. Nach diesem Ergebnis wäre von einer geringeren Diagnostizität der RSME-Skala für mentale Beanspruchung auszugehen als ursprünglich angenommen. Hiermit lässt sich auch der geringe Zusammenhang zwischen Modell und RSME über die vier Szenarien erklären. Die beschriebenen Zusammenhänge zwischen den einzelnen Messskalen konnten auch durch Korrelationen bestätigt werden. Somit konnte gezeigt werden, dass das Modell diagnostisch für mentale, nicht aber andere Beanspruchungsarten wie beispielsweise motorische Beanspruchung ist, was Ziel der Modellentwicklung war.

6.3.8 Überprüfung der Modellgüte mittels Einzelfallanalysen (Hypothese 9)

Bei der Überprüfung der Modellgüte mittels Korrelationen der Einzelfälle ergab sich das eindeutige Bild, dass ein positiver Zusammenhang zwischen den Modelldaten und den subjektiven Bewertungen der mentalen Beanspruchung mittels RSME innerhalb der Szenarien

besteht. Es ergaben sich wenige negative Korrelationen und Nullkorrelationen (5 bzw. 6). In den grafischen Darstellungen aller Versuchspersonen mit negativer Korrelation ist zu erkennen, dass die subjektiven Bewertungen dieser Versuchspersonen über die Zeit niedriger ausfallen. Es kann angenommen werden, dass hierbei der Anker für die subjektive Bewertung zunächst höher angesetzt wurde und im Verlauf des Szenarios ein Lern- und Gewöhnungseffekt einsetzten, so dass die Versuchspersonen ihre Bewertungen nach unten korrigierten. Derartige Effekte werden über das Modell nicht abgebildet. Zusätzlich könnte auch hier durch das Vorsortieren der Flugstreifen in der Stripbay ein Priming der Versuchspersonen stattgefunden haben, wodurch aufgrund der Erwartung höherer Belastung in den ersten Minuten der Szenarien mit hoher Belastung höhere Bewertungen vorgenommen wurden als in den Szenarien mit niedriger Belastung, weil eine höhere Anzahl vorsortierter Flugstreifen eine höhere Belastung vermuten lässt. Ebenso könnte gegen Ende des Szenarios die Erwartung, dass das Szenario bald zu Ende ist, ein Absinken der subjektiven Beanspruchungsbewertung zur Folge haben.

Weiterhin fiel allerdings auf, dass zwei der Versuchspersonen, für die sich negative Korrelationen bzw. Nullkorrelationen ergaben, über alle Szenarien hinweg und für fast alle Messzeitpunkte das Kreuz für ihre Bewertung auf ein- und demselben verbalen Anker der RSME-Skala setzten, so dass eine sehr geringe Variabilität der Messwerte zustande kommt. Eine Erklärung hierfür könnte sein, dass die Versuchspersonen eine Strategie anwendeten, in der sie der Bewertung als Nebenaufgabe eine sehr geringe Priorität zuordneten, um der Hauptaufgabe so viel Aufmerksamkeit wie möglich widmen zu können. Die Versuchspersonen könnten allerdings auch tatsächlich keinen Unterschied in der mentalen Beanspruchung empfunden haben. Ebenso könnte es daran liegen, dass die Versuchspersonen nicht ausreichend motiviert waren, um wirklich valide Bewertungen ihrer mentalen Beanspruchung abzugeben.

Bei den visuellen Vergleichen der Verläufe fiel auf, dass die Bewertungen mittels RSME und die Modelldaten zum Ende des Szenarios teilweise auseinanderdriften. Auch dies kann mit der Erwartung der Versuchspersonen erklärt werden, dass das Szenario bald zu Ende ist, wodurch die Bewertungen nach unten korrigiert werden. Ferner können aber auch Unterschiede in der Abarbeitung der Flugzeuge von Modell und Versuchspersonen dafür verantwortlich sein. Hierbei kann es passieren, dass die Versuchspersonen bestimmte Flugzeuge nicht oder anders abarbeiten als das Modell (d.h. die Versuchspersonen Flugzeuge entweder warten lassen, sie ignorieren oder es durch Fehler zu einer veränderten Belastung kommt, z.B. wenn ein Flugzeug durchstarten muss oder Freigaben zu früh/zu spät gegeben werden), was sich auf den Ablauf des weiteren Szenarios auswirkt und die Belastung verringert bzw. verändert (Folgeanfragen treten beispielsweise später oder früher auf als im Modell). Allgemein kann davon ausgegangen werden, dass bei einer Standardisierung der Szenarien und der damit verbundenen Möglichkeit zur Aggregation der Daten eine höhere Korrelation und ein besserer Fit der Kurven zu erwarten

wäre, da Ausreißer und individuelle Unterschiede nicht so stark ins Gewicht fallen. Letzten Endes wäre aber auch bei einem Einsatz des Modells in der Realität davon auszugehen, dass die Beanspruchung von Einzelpersonen und nicht die eines Personenkollektives im Fokus des Interesses steht, so dass das hier getätigte Vorgehen als sehr viel realitätsnaher anzusehen ist und die erhaltene Höhe der Korrelationskoeffizienten vor diesem Hintergrund als sehr gut einzuschätzen ist.

Für die Höhe der Korrelationen könnte allerdings eine mögliche geringe Diagnostizität der RSME-Skala für mentale Beanspruchung ein Problem dargestellt haben, wie sie die Ergebnisse der mittleren Beanspruchungshöhe (vgl. Kapitel 6.3.7) vermuten lassen. Sollten bei den Bewertungen mittels RSME-Skala tatsächlich andere Beanspruchungsarten bzw. Faktoren, welche die wahrgenommene Beanspruchung beeinflussen, mit abgebildet werden, könnte auch dies ein Grund für die relativ geringen Korrelationen sein. Die mit der RSME-Skala erfassten Daten würden somit nicht nur die reine mentale Beanspruchung wiedergeben sondern eher einen Gesamteindruck der Beanspruchung, sodass der Zusammenhang mit dem Modell, welches ausschließlich mentale Beanspruchung widerspiegelt, geringer ist, als wenn das Maß diagnostisch für mentale Beanspruchung ist.

Ebenso bestand die Schwierigkeit, dass beide Messskalen (also die des RSME und des Modells) relativ abstrakt und willkürlich festgelegt sind. Eine 1:1 Umsetzung der Modellskala zur Erfassung der mentalen Beanspruchung der Versuchspersonen wäre sinnvoll. Dabei könnte einerseits das Instantanious Self Assessment (ISA; Tattersall & Foord, 1996) als Vorbild dienen (z.B. auch durch eine Erweiterung der ISA-Skala von 5 auf 10 Punkte). Ebenso könnte die Sub-Skala „geistige Beanspruchung“ des NASA-TLX als Erfassungsskala während der Szenarien verwendet werden (die NASA-TLX-Skala ist zehnstufig). Hierfür spricht u.a. dass gute Korrelationen über die Szenarien hinweg erzielt wurden. Auch kann eine höhere Diagnostizität für mentale Beanspruchung angenommen werden, dies allerdings nur unter der Voraussetzung, dass auch die anderen Sub-Skalen ausgefüllt werden. Trotzdem besteht bei nahezu allen subjektiven Verfahren das Problem der nicht gegebenen Äquidistanz der Skala.

Die Anwendung des ISA oder der Durchführung der subjektiven Bewertungen am Rechner könnte zusätzlich das Problem der fehlenden Ratings lösen (vgl. S. 175). Mit diesem vorgehen wäre erfassbar, zu welchem Zeitpunkt eine Bewertung ausgelassen wurde und damit zu welchem Zeitpunkt der jeweilige Wert fehlt. So können auch bei fehlenden Werten die zeitlichen Verläufe erstellt sowie die Korrelationskoeffizienten berechnet werden.

Eine analoge Betrachtung der Korrelation zwischen Herzrate und Modelldaten (Hypothese 10) konnte aufgrund der in Hypothese 2 festgestellten geringen Sensitivität nicht erfolgen.

6.3.9 Explorative Analyse: Absolute Abweichung der Modelldaten

Für die Modelldaten wurde festgestellt, dass diese die mittlere Beanspruchung im Vergleich zu den subjektiven Daten, vor allem bei geringer Belastung, tendenziell eher unterschätzen. Sie spiegeln allerdings annähernd die Werte der RSME-Skala wider, was durch eine geringe mittlere absolute Abweichung belegt werden konnte. Die durch die RSME-Skala und das Modell erhobene mittlere Beanspruchungshöhe ist ferner meist nur halb so hoch, wie die des NASA-TLX g. A. Dies kann, wie bereits in Kapitel 6.3.4 erläutert, einerseits durch die Instruktion, andererseits durch den unterschiedlichen Erhebungszeitpunkt des NASA-TLX erklärt werden. Hierbei kann wiederum die Berechnungsmethode der mittleren Beanspruchungshöhe eine Rolle spielen. Wie beim RSME wird auch beim Modell die mittlere Beanspruchungshöhe aus den zu den 15 Messzeitpunkten ausgelesenen Beanspruchungshöhen errechnet. Hierbei gehen geringe und hohe Beanspruchung gleichermaßen ein. Für den NASA-TLX gilt allerdings analog zu oben die Annahme, dass bei der Bewertung des Gesamtszenarios Zeiträume höherer Belastung stärker in das Gesamturteil einfließen als Zeiträume geringerer Belastung, sodass höhere Beanspruchungsbewertungen zustande kommen. Ebenso können Unterschiede in den Skalen dafür verantwortlich sein, dass die Versuchspersonen mit der RSME-Skala ihre mentale Beanspruchung besser einschätzen können (vgl. Kapitel 6.3.4).

Es ist allerdings schwierig zu beurteilen, ob die Versuchspersonen bei Abgabe der RSME-Bewertung tatsächlich ihre aktuell empfundene Beanspruchung einschätzten, oder ob sie nicht doch eher einen Eindruck der Beanspruchung der gesamten vergangenen Minute abgaben. Dabei kann wiederum angenommen werden, dass hohe Beanspruchung innerhalb einer Minute stärker in die Bewertung mit einfließt als niedrige.

Allgemein ist anzunehmen, dass eine weitere systematische Verzerrung die mittlere subjektive Beanspruchungsbewertung der Versuchspersonen beeinflusste: So kann durch das Vorsortieren der Flugstreifen in der Stripbay ein Priming der Versuchspersonen stattgefunden haben, wodurch anhand der Anzahl der Flugstreifen eine höhere Bewertung der Gesamtbeanspruchung mittels NASA-TLX impliziert worden sein könnte.

6.3.10 Explorative Analyse: Fehler und Beanspruchungspeaks

Zunächst ist darauf hinzuweisen, dass diese Analyse keine Standardmethodik darstellt, sondern lediglich der Sichtbarmachung der oben getätigten Annahme dienen soll, dass Fehler vor allem zu Zeiten hoher Belastung und Beanspruchung wahrscheinlicher sind. Es gilt zu berücksichtigen, dass ein Beanspruchungsgipfel des Modells nicht zwangsläufig mit einem Fehler der Versuchspersonen zum selben Zeitpunkt einhergehen muss, sondern sich lediglich seine Auftretenswahrscheinlichkeit erhöht, wodurch es schwierig ist falsche Alarme als auch korrekte Zurückweisungen zu beurteilen.

Es konnte gezeigt werden, dass 61 % der von den Versuchspersonen gemachten Fehler innerhalb von Beanspruchungspeaks lagen, die auch das Modell vorhersagte. Wird also ein Beanspruchungspeak vom Modell für ein bestimmtes Szenario vorhergesagt, so kann mit 61 prozentiger Sicherheit davon ausgegangen werden, dass auch ein Fehler innerhalb dieses Peaks auftreten wird. 183 der insgesamt 300 Fehler konnten vom Modell vorhergesagt werden.

Problematisch bei dieser Art der Beschreibung ist, dass nur Treffer eindeutig beschrieben werden können. Es können einerseits ebenso Fehler zu Zeiten geringer Beanspruchung auftreten (was zu Auslassungen führen würde), sowie auch in Beanspruchungspeaks innerhalb derer keine Fehler beobachtet wurden (falsche Alarme), die Wahrscheinlichkeit für Fehler erhöht gewesen sein. Daher ist das vorliegende Ergebnis schwierig zu interpretieren. Andererseits ist für die Beurteilung der Modellgüte prinzipiell auch zu beachten, inwieweit zu Zeitpunkten ohne Beanspruchungspeaks auch keine Fehler auftreten (korrekte Zurückweisungen). Es ist davon auszugehen, dass Fehler auch auftreten können, wenn geringe Belastung und/oder Beanspruchung vorliegen, z.B. durch Unterforderung. Solche Zusammenhänge sind im Modell bisher nicht implementiert und können daher nicht abgebildet werden. Auch ist das Modell nicht explizit darauf ausgelegt Fehler vorherzusagen oder Fehler zu machen. Der Informationsgehalt des Anteils an Treffern und Auslassungen ist daher kritisch zu beurteilen. Dennoch kann anhand der Ergebnisse festgehalten werden, dass das Modell prinzipiell auch in der Lage ist, Fehler vorherzusagen und damit ein Indiz für die Güte des Modells auch auf der Mikroebene gegeben werden kann. Der erreichte Wert liegt zwar nur 11 % über der Ratewahrscheinlichkeit von 50 %, allerdings ist dieses Ergebnis, vor dem Hintergrund, dass nur Treffer eindeutig beschrieben werden können, als gut zu beurteilen.

6.4 Kapitelzusammenfassung

In der Studie zur Validierung des erstellten Modells wurden Beanspruchungsdaten von 24 Versuchspersonen in einer Towersimulation erhoben und mit den Modelldaten verglichen. Jede Versuchsperson hatte dabei für vier 15-minütige Verkehrsszenarien die Aufgaben eines Towerlotsen an einem kleinen Flughafen zu übernehmen, also Landungen und Starts, sowie das Rollen von Flugzeugen durch die Erteilung der richtigen Freigaben zu koordinieren (vgl. Kapitel 6.1). Zunächst konnte kein Einfluss des Abschneidens im d2-R-Test und damit des Effizienz-Gründlichkeits-Tradeoff auf die Fehlerzahl und die subjektive Beanspruchungsbewertung festgestellt werden (vgl. Kapitel 6.2.1).

Die Auswertung der Herzrate ergab, dass deren Sensitivität nicht ausreichte, um zwischen verschiedenen Belastungsbedingungen zu unterscheiden. Es konnten aber Unterschiede zwischen den beiden Experimentalbedingungen und der Baseline-Bedingung nachgewiesen werden. Hierdurch wurden Befunde von Nickel et al. (2002) bestätigt, dass die Herzrate nicht

sensitiv genug ist, um *„feiner*[72] *abgestufte Belastungsgrade“* (S. 33) aufzudecken, sondern lediglich in der Lage ist, eine Experimental- von einer Baseline-Bedingung zu unterscheiden (vgl. Kapitel 6.2.2). Hierdurch erweist sich die Herzrate für weitere Untersuchungen als nicht sinnvoll.

Die statistische Auswertung erfolgte zunächst hinsichtlich der experimentellen Manipulation. Es konnte festgestellt werden, dass die durch subjektive Methoden erfasste mentale Beanspruchung in den meisten Fällen auf die höhere Anzahl Flugzeuge zurückgeführt werden konnte, da bei einer größeren Anzahl Flugzeuge die Beanspruchung signifikant höher eingeschätzt wurde, als bei einer geringeren Anzahl Flugzeuge. Zusätzlich unterschied sich die Beanspruchungshöhe bei gleicher Belastung in der Mehrzahl der Fälle nicht signifikant. Nur vereinzelt konnten auch Unterschiede bei gleicher Belastung festgestellt werden, was auf eine unterschiedliche Verteilung der Flugzeuge über die Zeit und damit eventuell zustande kommende unterschiedliche schwierige Verkehrssituationen zurückgeführt werden kann (Kapitel 6.2.4 und 6.2.5). Auch die Ergebnisse der Auswertung der Fehlerzahl können dies untermauern (Kapitel 6.2.6).

Durch die Analysen der mittleren Beanspruchungshöhe der subjektiven Maße sowie der mittleren Fehlerzahl konnte eine erfolgreiche experimentelle Manipulation nachgewiesen werden. Es wurde festgestellt, dass Szenarien mit unterschiedlicher Belastung zu signifikant unterschiedlichen Bewertungen führen. Szenarien mit gleicher Belastung führen zu keinen signifikanten Unterschieden in den Bewertungen. Ein analoger Hypothesentest der Modelldaten konnte dieses Bild nicht vollständig replizieren. Hier wurden Unterschiede zwischen allen Szenarien (also auch zwischen denen mit gleicher Belastung) signifikant. Die Ursache wird in der geringen Varianz der Modelldaten gesehen, wodurch die Varianzanalyse bereits bei kleinen Mittelwertsunterschieden signifikante Ergebnisse lieferte (vgl. Kapitel 6.2.7 und 6.3.6).

Im Trend der Mittleren Beanspruchungsratings über alle vier Szenarien konnte festgestellt werden, dass das Modell nahezu exakt den Trend, der mit dem NASA-TLX g. A. erhoben wurde, abbildet. Die Trends des NASA-RTLX sowie des RSME werden schlechter vom Modell repliziert. Hierbei wird vermutet, dass mittels RSME nicht ausschließlich mentale Beanspruchung erfasst wird, sondern andere Beanspruchungsarten und Faktoren, die die wahrgenommene Beanspruchung der Versuchspersonen beeinflussen, mit in die Bewertung einfließen, da hier die Korrelation zum NASA-RTLX höher ist als zum NASA-TLX g. A. (vgl. Kapitel 6.2.8 und 6.3.7). Die gute Übereinstimmung des Modells mit dem NASA-TLX g. A. kann als Beleg für die Diagnostizität des Modells für mentale Beanspruchung gesehen werden.

[72] In der vorliegenden Studie kann sogar davon ausgegangen werden, dass erhebliche Unterschiede in der Belastung bestanden (10 vs. 20 Flugzeuge pro 15 Minuten), wobei auch diese erheblichen Unterschiede mittels Herzrate nicht aufgedeckt werden konnten.

Zur weiteren Überprüfung der Modellgüte wurden Einzelfallanalysen für alle 96 Versuchspersonenszenarien durchgeführt (4 Szenarien x 24 Versuchspersonen) und die zeitlichen Verläufe der RSME-Erfassung mit den zugehörigen zeitlichen Verläufen der Modellsimulationen sowohl visuell als auch korrelativ verglichen. Einige Vergleiche wurden aufgrund fehlender Messwerte aus der Betrachtung ausgeschlossen. Hierbei konnte für die überwältigende Mehrzahl der Versuchspersonen ein positiver Zusammenhang zwischen Modelldaten und subjektiven Beanspruchungsdaten gezeigt, sowie im Mittel ein schwach bis mittlerer korrelativer Zusammenhang festgestellt werden (vgl. Kapitel 6.2.9 und 6.3.8).

Der explorative Vergleich der mittleren Beanspruchung aus den Modelldaten mit den subjektiven Beanspruchungsdaten zeigt ferner, dass das Modell die absolute Höhe der mittleren Beanspruchung des NASA-TLX g. A. meist um etwa den Faktor 2 unterschätzt. Die absolute Höhe der Modelldaten stimmt aber gut mit denen des RSME überein (MAD = 5,36 %; vgl. Kapitel 6.2.10 und 6.3.9).

Die weitere explorative Analyse ergab außerdem, dass gut 61 % der von den Versuchspersonen gemachten Fehler innerhalb von Beanspruchungspeaks des Modells lagen (vgl. Kapitel 6.2.11), wodurch das Modell prinzipiell auch in der Lage ist, Fehler vorherzusagen und ein Indiz für die Gültigkeit des Modells auf der Mikroebene geschaffen werden konnte (vgl. Kapitel 6.3.10).

7 Diskussion der Arbeit

Im Folgenden wird die Arbeit anhand der in Kapitel 1.3 gesetzten Forschungsziele diskutiert.

7.1 Beitrag zur Ableitung einer geeigneten Modellierungsform des Konstruktes mentaler Beanspruchung (Forschungsziel 1)

Das Forschungsziel der Ableitung einer geeigneten Modellierungsform des Konstruktes mentaler Beanspruchung kann anhand der im Kapitel 2 vorgestellten Theorien bewertet werden. Hierbei wurde zunächst analysiert, welche theoretischen Vorstellungen von mentaler Beanspruchung existieren. Zwei einschlägige Theorien, die sich mit dem theoretischen Konstrukt befassen, wurden vorgestellt, wobei für die vorliegende Arbeit insbesondere die aufmerksamkeitstheoretischen Modelle als Basis für die Modellierung mentaler Beanspruchung verwendet wurden. Auch wenn die meisten Definitionen mentaler Beanspruchung aufmerksamkeitstheoretische Modelle zugrunde legen, so können sie doch nicht erklären, warum mentale Beanspruchung durch die Ableitung physiologischer Indikatoren erfassbar sein soll. Dies gelingt erst durch die Integration aktivierungs- und aufmerksamkeitstheoretischer Modelle.

Da aus der Untersuchung der Theorien aufmerksamkeitstheoretischer Modelle hervorgeht, dass vor allem die Begrenzungen des Arbeitsgedächtnisses eine entscheidende Rolle spielt, wurde diese, als für das basale Modell zunächst wichtigster Einflussfaktor mentaler Beanspruchung, in das Modell übernommen. Hierbei wurde die mentale Beanspruchung über die momentan vorhandene Anzahl der Chunks im Arbeitsgedächtnis abgebildet. Im Gegensatz zu Annahmen über das Arbeitsgedächtnis, wie sie Theorien der Produktionensysteme tätigen, die keine explizite Arbeitsgedächtnisstruktur annehmen, sondern das Arbeitsgedächtnis als aktiven Teil des Langzeitgedächtnisses ansehen, wird hier die Ansicht vertreten, dass Operateure zur Aufgabenbearbeitung Informationen im Arbeitsgedächtnis aufrechterhalten müssen. Es wurde eine Single-Limited-Capacity Theorie des Arbeitsgedächtnisses verwendet, die eine einzelne Ressource mit einer Obergrenze beschreibt. Zwar sprechen neuere Untersuchungsergebnisse dafür, dass auch das Arbeitsgedächtnis über multiple Ressourcen verfügt. Es existieren verschiedene Modellvorstellungen über den Aufbau des Arbeitsgedächtnisses, die alle experimentell sehr gut belegbar sind. Allerdings ist es schwierig zu sagen, welche Theorie die richtige oder zumindest „richtiger“ ist. Ferner existieren, obwohl die Untersuchung des Aufbaus und der Funktion der einzelnen Strukturen des Arbeitsgedächtnisses sehr weit fortgeschritten ist, bis heute wenige gesicherte Erkenntnisse darüber, wie die einzelnen Strukturen in ihrer Kapazität begrenzt sind. Die konkretesten und wohl auch naheliegendsten Annahmen hierzu macht Cowan (1998), der eine zeitliche, weniger eine absolute Begrenzung des

Arbeitsgedächtnisses nahelegt, wie sie auch in der kognitiven Architektur ACT-R implementiert ist. Die Annahme einer solchen zeitlichen Begrenzung ist allerdings wiederum an Theorien der Produktionensysteme angelehnt. Ferner ist eine solche Begrenzung im ereignisdiskreten Modellierungssystem Petrinetz sehr schwer umzusetzen (vgl. Kapitel 5.10). Cowan (2001) beschreibt weiterhin auch eine absolute Begrenzung von vier Chunks, welche im Fokus of Attention liegt. Hierbei ist unklar, wie diese Befunde einer Kapazität von vier Chunks (Oberauer und Bialkova (2009) sprechen sogar nur von einem Chunk) mit anderen Befunden von Miller (1956) mit fünf bis neun Chunks oder Bisseret (1971) und Sperandio (1969) mit zehn Chunks bei Lotsen in Einklang zu bringen sind. Um die Quantifizierung mentaler Beanspruchung zu gewährleisten, erweist sich daher die einfache Annahme einer Single-Limited-Capacity Theorie des Arbeitsgedächtnisses als sehr nützlich.

Aus dem zur Validierung des Modells durchgeführten Experiment lässt sich nur schwer sagen, welche maximale Arbeitsgedächtniskapazität vorliegt, da kein Test der Arbeitsgedächtnis-kapazität durchgeführt wurde. Anzunehmen ist aber, dass das Auftreten von Fehlern im Experiment mit einem Erreichen der Obergrenze der speicherbaren Chunks zusammenhängt. So können bei erreichen der Obergrenze nicht mehr alle nötigen Informationen ausreichend repräsentiert werden, weshalb schlechtere Entscheidungen getroffen werden. Die Annahme einer maximalen Kapazität von zehn Chunks ist dabei kritisch zu reflektieren. Da das Modell die Performanz von Experten abbilden soll, ist es sinnvoll sich an der maximalen Kapazität zu orientieren, die bei Experten angenommen wird. Es ist aber einerseits anzunehmen, dass auch bei Experten diese Obergrenze nach oben und unten schwankt (z.B. 10 ± 2 Chunks), da es sich um einen Mittelwert handelt. Andererseits ist fraglich, ob diese Obergrenze auch im getätigten Validierungsversuch angenommen werden kann, da diese mit einer Laienstichprobe durchgeführt wurde. So ist anzunehmen, dass die Obergrenze bei den Versuchspersonen schon früher erreicht wird, als im Modell. Es wurde jedoch versucht, die Aufgabe recht einfach zu gestalten und den Versuchspersonen einen gewisser Grad an Expertise zu übermitteln.

Auch wenn für das Arbeitsgedächtnis eine Single-Limited-Capacity Theorie implementiert wurde, so sind aber auch Ansätze multipler Ressourcen in das Modell eingeflossen. Visuelle und auditive Aufmerksamkeitsressourcen, sowie die motorische Handlungsausführung sind beispielsweise als vom Arbeitsgedächtnis relativ unabhängige Ressourcen modelliert. Diese fließen zwar nicht direkt in die Quantifizierung mentaler Beanspruchung ein, sorgen allerdings für eine relativ serielle Aufnahme von Chunks in das Arbeitsgedächtnis und Abarbeitung motorischer Aufgaben und berücksichtigen die dafür benötigte Zeit. Diese Betrachtung gewährleistet, dass visuelle Aufmerksamkeit jeweils nur auf einer bestimmten Informationsquelle liegen kann, bzw. nur eine motorische Handlung gleichzeitig ausgeführt werden kann.

Wenig geklärt ist dabei allerdings, welche Rolle Begrenzungen des perzeptiven Apparates bei der subjektiven Bewertung mentaler Beanspruchung spielen und wie diese das empfinden mentaler Beanspruchung beeinflussen können. Es ist damit sehr schwierig zu bestimmen, wie die einzelnen Phasen der Informationsverarbeitung zusammenwirken und das Empfinden mentaler Beanspruchung hervorbringen. Hierbei ist weitere intensive Grundlagenforschung nötig, um diese Fragen beantworten zu können und eine Möglichkeit zu finden, auch die Prozesse der Informationsaufnahme, in die Bestimmung eines Beanspruchungsindex mit aufzunehmen.

Insgesamt wird der Ansatz, die mentale Beanspruchung in Abhängigkeit des Verhältnisses von maximal verfügbaren zu bereits verbrauchten Informationsverarbeitungsressourcen (Anzahl Chunks im Arbeitsgedächtnis) zu modellieren, als sehr vielversprechend angesehen. Es konnte ein theoretisch fundierter und psychologisch plausibler Ansatz gefunden werden mentale Beanspruchung zu quantifizieren, der so bisher nicht vorlag.

7.2 Beitrag zur Verfolgung eines makrokognitiven Modellierungsansatzes und Verbindung mit mikrokognitiven Ansätzen (Forschungsziel 2)

In der vorliegenden Arbeit wurde die Verbindung mikro- und makrokognitiver Ansätze angestrebt um Nachteile beider auszugleichen und Vorteile nutzen zu können. Hierbei sind reine makrokognitive und rein mikrokognitive Ansätze, sowie die Anwendung des *Downscaling* zur Einbettung mikrokognitiver Funktionen in ein makrokognitives Modell zu diskutieren.

Der in dieser Arbeit verfolgte *makrokognitive Ansatz* kann vor allem für stark anwendungsbezogene Problemstellungen als angemessen angesehen werden. Durch die anwendungsnahe Betrachtung des Mensch-Maschine-Systems Flugverkehrskontrolle können mit Hilfe des Modells ökologisch sehr valide Vorhersagen gemacht werden. Die recht grobe Betrachtungsweise kognitiver Prozesse ist hierbei allerdings kritisch zu diskutieren. Die makrokognitive Betrachtungsweise hat zum Ziel, Modelle zu erstellen, die dasselbe Verhalten zeigen und dieselbe Performanz erbringen, wie es der Mensch tun würde. Dabei werden allerdings kognitive Prozesse nicht im Detail betrachtet, was im Gegensatz zu kognitiven Architekturen steht. Es gilt darum, zu beachten, dass durch das „Weglassen“ von Details im Hinblick auf kognitive Prozesse keine „Black-Box-Modelle“ entstehen. Prozesse sollten daher so detailliert wie nötig betrachtet werden, um eine große Korrespondenz zwischen Modell und Wirklichkeit gewährleisten zu können. Probleme, die die Entstehung der menschlichen Performanz und des menschlichen Verhaltens auf Grundlage kognitiver Prozesse untersuchen sollen, können daher mit rein makrokognitiven Betrachtungsweisen nicht untersucht werden.

Rein mikrokognitive Betrachtungsweisen hingegen, können das Verständnis über die Entstehung kognitiver Prozesse und menschlichen Verhaltens fördern, die Verknüpfung zu

anwendungsorientierten Problemstellungen gestaltet sich allerdings meist schwierig. Die Gefahr besteht hierbei darin, dass immer kleinere Problembereiche erforscht werden (wir also *„immer mehr über immer weniger wissen";* Kindsmüller et al., 2004, S. 10) und der Anwendungsbezug damit noch stärker verloren geht.

Da die Unterscheidung in Mikro- und Makrokognition keine Schwarz-weiß-Kategorien darstellt, sondern beide eher als die zwei Endpunkte eines Kontinuums anzusehen sind, wurde für diese Arbeit an einigen Stellen durch *Downscaling* eine Verbindung zwischen makrokognitiver und mikrokognitiver Modellierung vorgenommen. So wurde das Modell zwar stark anwendungsorientiert auf die Aufgaben von Lotsen im Flughafentower ausgelegt, aber dennoch zahlreiche kognitionspsychologische Theorien der Informationsverarbeitung in die Beschreibung der kognitiven Prozesse des modellierten Lotsen einbezogen. So orientiert sich beispielsweise das Wahrnehmungsmodell vor allem am Multiple-Ressource-Modell von Wickens (1984).

Dennoch ist, wie auch die recht grobe makrokognitive Betrachtungsweise kognitiver Prozesse, die Übertragung mikrokognitiver Theorien zur Beschreibung von kognitiven Vorgängen in einer Anwendungsdomäne kritisch zu reflektieren. Die Annahmen einer Theorie werden vor allem durch Laborexperimente, meist in Isolation und unter starker Kontrolle mit relativ homogenen Stichproben (meist Studenten) beobachtet. Hierbei kann es möglich sein, dass Verhalten beobachtet und Prozesse entdeckt werden, welche sich in einer natürlichen Umgebung und unter natürlichen Bedingungen mit Experten als Akteuren nicht zeigen würden oder eine verschwindend kleine Rolle spielen.

Durch die Verbindung von Theorie und Anwendungsdomäne wurde versucht, die im Fokus stehenden Prozesse hinreichend detailliert abzubilden, wodurch ein Modell im Sinne des „Minimal Modelling Manifesto" erhalten werden kann (Hollnagel, 1993b, S. 379 vgl. auch Hollnagel, 1993a): *„A Minimal Model is a representation of the main principles of control and regulation that are established for a domain – as well as for the capabilities and limitations of the controlling system."*

Im Hinblick auf eine weitere Verbesserung der modellhaften Beschreibung mentaler Beanspruchung sind weitere *Downscaling*-Prozesse notwendig, um beispielsweise die zur Entscheidungsfindung von Lotsen nötigen Prozesse noch adäquater abbilden zu können. So ging bereits aus der Aufgabenanalyse hervor, dass vor allem Entscheidungen, die mit dem Rollen von Flugzeugen zutun haben, hochgradig auf mentalen Simulationsprozessen beruhen, welche dazu dienen, zukünftige Verkehrssituationen vorauszusehen. Die im vorliegenden Modell umgesetzte Entscheidungsfindung auf Grundlage der aktuellen Verkehrssituation, wird daher weder der Komplexität der Aufgabe, noch der damit verbundenen kognitiven Prozesse gerecht. Die

Schwierigkeiten, die die Implementierung detaillierterer Entscheidungsprozesse für die Modellierung mit sich bringt, beschreiben Smieszek und Joeres (2013), weshalb auf Grund fehlender Grundlagenforschung in der vorliegenden Modellversion vereinfachte Annahmen getroffen werden mussten.

Die Vorgehensweise des *Downscaling* kann alternativ durch die Annäherung von der mikrokognitiven Ebene zur makrokognitiven Ebene (*Upscaling*) erfolgen. So konnten West et al. (2013) zeigen, dass es möglich ist, makrokognitive Funktionen mit Hilfe einer mikrokognitiven Architektur zu beschreiben. Dies beschränkt sich allerdings bisher auf wenige makrokognitive Funktionen, die auch in Isolation betrachtet wurden. Weitere Forschung ist notwendig, um hier ein adäquates *Upscaling* möglich zu machen.

In der vorliegenden Arbeit kann das *Downscaling* und damit die Verbindung von mikro- und makrokognitiver Modellierung als gelungen angesehen werden. Kognitive Prozesse wurden soweit möglich und nützlich für die Problemstellung auf ein Mikroniveau heruntergebrochen, andere Prozesse die für die vorliegende Problemstellung weniger wichtig sind wurden klassisch makrokognitiv betrachtet. Aber auch durch die Identifizierung weiteren Forschungspotentiales innerhalb der Grundlagenforschung konnte eine Verbindung zwischen makro- und mikrokognitiven Ansätzen geschaffen werden. Die Entwicklung des makrokognitiven Modells MATriCS hat dazu beigetragen, dass Forschungspotential identifiziert wurde, welches zunächst mikrokognitiv zu untersuchen ist.

7.3 Beitrag hinsichtlich der Erstellung eines Modells zur quantitativen Beschreibung mentaler Beanspruchung von Towerlotsen in Abhängigkeit verschiedener Einflussfaktoren (Forschungsziel 3)

Das Forschungsziel der Erstellung eines Modells zur quantitativen Beschreibung mentaler Beanspruchung von Towerlotsen in Abhängigkeit verschiedener Einflussfaktoren ist zunächst hinsichtlich der Umsetzung der identifizierten Theorien mentaler Beanspruchung, sowie der Studien zu Einflussfaktoren mentaler Beanspruchung zu diskutieren. Ferner sind getätigte Vereinfachungen des Modells kritisch zu reflektieren. Ebenso ist die Verwendung Farbiger Petrinetze als Beschreibungsmittel zu bewerten.

Es konnte gezeigt werden, dass mit Hilfe des erstellten Modells die mentale Beanspruchung von Towerlotsen über die Anzahl der im Arbeitsgedächtnis gespeicherten Chunks operationalisiert und sichtbar gemacht werden kann. Durch das Auslesen der Anzahl der Chunks aus dem Modell können kontinuierliche Verläufe der Arbeitsgedächtnisbelastung von Lotsen bei der Arbeit erhalten werden. Hierbei wurde im vorliegenden Modell der wichtigste Beanspruchungsfaktor, die Anzahl der zu bearbeitenden Flugzeuge (vgl. Koros et al., 2003) verwendet. Nichtsdestotrotz stellt dieser Faktor nur einen von vielen Faktoren dar, die die

mentale Beanspruchung beeinflussen. Neben einer Vielzahl objektiver Faktoren, wie beispielsweise das „Kreuzen der aktiven Runway“, „Überlastungen der Frequenz“, „unterschiedliche Leistungscharakteristika von Flugzeugen“, „Sichtbedingungen“ (vgl. Koros et al., 2003; Kapitel 4.2) spielen auch subjektive Faktoren eine Rolle, wovon in dieser Arbeit nur die maximale Arbeitsgedächtniskapazität mit einbezogen wurde. Viele weitere subjektive Faktoren können allerdings einen Einfluss auf das subjektive Beanspruchungsempfinden besitzen, wie beispielsweise die Motivation, das Training, kognitiven Strategien oder Regulations-mechanismen mentaler Beanspruchung. Sperandio (1978) beschreibt beispielsweise, dass sich Lotsen über ihre Beanspruchungshöhe durchaus bewusst sind und somit bei drohender Überlastung ihre Strategie anpassen, um dieser Überlastung entgegenzuwirken (vgl. auch Loft et al., 2007). Eine solche Strategieanpassung kann beispielsweise das einfache Verschieben von Flügen sein auf einen späteren Zeitpunkt sein (vgl. Smieszek et al., 2013). Dennoch ist auch das Bewusstsein des Lotsen über seine eigene Beanspruchung kein Garant dafür, dass rechtzeitig eine Strategieanpassung vorgenommen wird. Es kann bereits zu spät sein, wenn die Strategieanpassung erfolgt. Ferner ist anzunehmen, dass bei einer Strategieanpassung eher an Gründlichkeit eingespart wird (Verwendung von Short-Cuts), um das Effizienzziel aufrechtzuerhalten, als das ein low risk action regulation mechanism angewendet wird.

Ebenso bleibt im Modell unberücksichtigt, wie der Anstieg der mentalen Beanspruchung die Güte der Entscheidungsfindung beeinflussen kann. Zwar sind einige der genannten Variablen bereits im Modell implementiert worden (z.B. „Belastung der Funkfrequenz“ oder „Sichtbedingungen“) – diese sind aber bei der Validierung des Modells nicht berücksichtigt worden, um einen adäquaten experimentellen Rahmen zu schaffen. Dennoch spielt der Einfluss all dieser Variablen eine mehr oder weniger entscheidende Rolle bei der Entstehung mentaler Beanspruchung. Zusätzlich kann auch durch die Interaktion zwischen diesen Variablen ein erheblicher Einfluss auf die mentale Beanspruchung entstehen. Durch die genauere Betrachtung, vor allem subjektiver Variablen im Modell, kann die Fähigkeit des Modells, individuelle Versuchspersonen-Beanspruchung abzubilden, erhöht werden.

Auch ist bereits erwähnt worden, dass der Prozess der Entscheidungsfindung recht einfach abgebildet worden ist. Hierbei wird lediglich die aktuelle Verkehrssituation zugrunde gelegt. Es wurde jedoch klar, dass mentale Simulationsprozesse vorgenommen werden, welche hochgradig die mentale Beanspruchung des Lotsen beeinflussen (Smieszek & Joeres, 2013). Diese müssen durch das Modell berücksichtigt werden. Hierzu sind zunächst grundlagenwissenschaftliche Fragen zu klären.

Eng damit verbunden ist auch die Umsetzung des Arbeitsgedächtnismodells. Smieszek und Joeres (2013) sehen als Folge eines veränderten, komplexeren Entscheidungsprozesses auch die Notwendigkeit eines veränderten Arbeitsgedächtnismodells. Dies legt auch die Existenz

zahlreicher Befunde nahe, wie es bereits in Kapitel 7.1 diskutiert worden ist. Die Umsetzung eines solchen Arbeitsgedächtnismodells erscheint prinzipiell sinnvoll, wobei bisher ungeklärt ist, wie Informationen in einem solchen Arbeitsgedächtnis repräsentiert werden (vgl. auch Kapitel 7.1).

Die Verwendung farbiger Petrinetze als Beschreibungsmittel zur Modellierung menschlichen Verhaltens und menschlicher Kognition kann als sehr vielversprechend, vor allem für die makrokognitive Modellierung bzw. für ein Downscaling von der makrokognitiven zur mikrokognitiven Ebene bezeichnet werden: So können kognitive Prozesse und Verhalten nicht nur abgebildet und simuliert, sondern auch sichtbar gemacht werden, wobei gleichzeitig die Komplexität der modellierten Systeme sehr gut beherrschbar ist. Gleichzeitig besteht, im Gegensatz zu kognitiven Architekturen, die Möglichkeit, dass Prozesse, die nicht im Fokus der Problemstellung stehen, vereinfacht dargestellt oder gar weggelassen werden können. Zusätzlich können auch bei im Fokus stehenden Prozessen zunächst vereinfachte Annahmen gemacht werden, um festzustellen, wie gut bereits einfache Modellannahmen in der Lage sind, experimentelle Daten abzubilden, so wie es beispielsweise beim Prozess der Entscheidungsfindung geschehen ist. Darüber hinaus ist die Möglichkeit die modellhafte Beschreibung des Mensch-Maschine-Systems auf ganzheitliche Weise vornehmen zu können von unschätzbarem Wert. Durch die Flexibilität des generischen Flughafenmodells, war es möglich, Schwächen, die die Simulationssoftware aufwies, ausgleichen zu können. Die Verwendung einer kognitiven Architektur hätte hierbei entweder die Anbindung an die vorhandene oder die Erstellung einer eigenen Simulationssoftware erfordert.

Die Verwendung eines ereignisdiskreten Modellierungswerkzeuges zur Beschreibung vielerlei kontinuierlicher Prozesse stellte sich hierbei allerdings als eher ungeeignet heraus (z.B. zur Modellierung der Rollprozesse am Flughafen oder des zeitlichen Verfalls von Informationen im Arbeitsgedächtnis, wie in Kapitel 5.4.10 beschrieben; vgl. Cowan, 1998). Hier könnten neuere Methoden, wie Real-Time Coloured Petri Nets (RTCP-nets; Szpyrka, 2006, 2008) oder Continuous und Hybrid Petri Nets (David & Alla, 2010) eine vielversprechende Alternative darstellen. Für diese Methoden fehlt es allerdings noch weitestgehend an Computer-Tools für die Erstellung lauffähiger Modelle, die von ihrer Stärke her mit dem für Coloured Petri Nets zur Verfügung stehenden CPN-Tools vergleichbar wären. Die grafische Darstellung und die damit verbundene Sichtbarmachung kognitiver Prozesse, die alle Arten von Petrinetzen mit sich bringen, birgt große Vorteile hinsichtlich der Transparenz von Modellen (vgl. Gore, Hooey, Foyle, & Scott-Nash, 2008). Ebenso ist die relativ intuitive Darstellung sehr hilfreich für die Verständigung und den Austausch mit anderen Wissenschaftlern.

Das Forschungsziel der Erstellung eines basalen Modells der Lotsenbeanspruchung kann damit als erreicht angesehen werden.

7.4 Beitrag hinsichtlich der Validierung des erstellten Modellansatzes mittels statistischer Methoden (Forschungsziel 4)

Neben der Erstellung des Modells war es ebenso wichtig, einen Nachweis darüber zu erbringen, dass das Modell in der Lage ist, adäquat die mentale Beanspruchung realer Versuchspersonen in realen Aufgabensituationen abzubilden. Hierbei kann vor allem die *Übertragbarkeit der Studie* durch die Verwendung einer Mid-Fidelity Simulation mit Laien als Versuchspersonen sowie die *Versuchsdurchführung* diskutiert werden. Hinsichtlich des Modell-Daten-Vergleiches können verschiedene Problematiken identifiziert werden. Es ist die *Schwierigkeit mit Hilfe der Simulation standardisierte Szenarien zu generieren* zu diskutieren, wodurch die Komplexität der Verkehrssituationen unterschiedlich ausfiel. Auf die *Beschaffenheit der subjektiven Skalen* sind vor allem Probleme hinsichtlich der *Berechnung von Abweichungsmaßen* aber auch die Notwendigkeit der *Anwendung nonparametrischer statistischer Verfahren* zurückzuführen. Ferner ist die *Sensitivität der Herzrate* zu beurteilen.

Zur Erhebung von Daten mentaler Beanspruchung für einen Modell-Daten-Vergleich wurde ein Simulationsexperiment im August 2013 durchgeführt. Das Ziel bestand vor allem darin, zu überprüfen, ob das Modell in der Lage ist die von Versuchspersonen in den Simulationsszenarien abgegebenen Bewertungen ihrer mentalen Beanspruchung zu replizieren.

Für die *Übertragbarkeit der Studie* und damit des Modells auf die Realität kann speziell die Verwendung einer Mid-Fidelity-Simulation und einer Laienstichprobe kritisch gesehen werden. Zwar wurde gewährleistet, dass die Simulation möglichst realistisch war, dennoch bestehen einige Unterschiede zu realen Arbeitssituationen: Beispielsweise haben echte Lotsen neben der Kommunikation mit den Piloten zahlreiche Absprachen mit Kollegen, Vorgesetzten und anderen Einrichtungen über das Telefon zu erledigen, die weder im Modellansatz noch im Experiment betrachtet wurden. Ebenso war die Aufgabe der Versuchspersonen recht einfach gehalten, da auf die Anfrage eines Flugzeuges nach Überblicken der Verkehrssituation lediglich mit der dazugehörigen Freigabe geantwortet werden musste. Die reale Kommunikation zwischen Lotse und Pilot gestaltet sich jedoch wesentlich komplexer. Diese Vereinfachungen waren einerseits der Komplexitätsreduktion im Modellansatz, andererseits der Verwendung von Laien als Versuchspersonen geschuldet. Zwar wurde gewährleistet, dass die Versuchspersonen durch ausführliche Instruktion und Training einige Erfahrung sammeln konnten, dennoch kann der Erfahrungsschatz echter Lotsen niemals innerhalb einer zweistündigen Sitzung an Laien übermittelt werden. Aufgrund dessen wäre auch eine maximale Arbeitsgedächtniskapazität von weniger als zehn Chunks anzunehmen (vgl. Bisseret, 1971; Kapitel 2.2 und 7.1). Für eine erste Validierungsstudie wird der Ansatz dennoch als zielführend angesehen, da so, mit relativ geringem Ressourceneinsatz, ein erster Nachweis der Güte des Modells erbracht werden konnte.

Nun gilt es im Weiteren, diese Befunde mit echten Lotsen in einer High-Fidelity-Simulation zu überprüfen.

Eine Schwierigkeit, welche die Simulation mit sich brachte, bestand darin, dass *Szenarien nicht standardisierbar* waren. Zwar konnten Zeiten definiert werden, zu denen Flugzeuge auftauchen sollten, die exakte Zeit variierte allerdings um diesen definierten Zeitpunkt herum. Durch die Flexibilität des Modells konnte zwar gewährleistet werden, dass das Modell immer genau das Szenario abarbeitete, was auch die Versuchsperson bearbeitet hatte – die Aggregation eines mittleren Verlaufes sowohl der Versuchspersonen- als auch der Modelldaten war dennoch nicht sinnvoll. Allgemein wird erwartet, dass bei standardisierten Szenarien die Versuchspersonen noch stärker in ihren Urteilen übereinstimmen, sodass die Korrelation von Versuchspersonenratings und Modelldaten höher ausfallen. Ferner ließe sich über die Aggregation der zeitlichen Verläufe ein mittlerer zeitlicher Verlauf erhalten. Hierbei wird einerseits erwartet, dass die Korrelation höher ausfallen wird, da individuelle Unterschiede und Ausreißer nicht so sehr ins Gewicht fallen, andererseits entsteht hierdurch die Möglichkeit, auch Abweichungsmaße für den zeitlichen Verlauf berechnen zu können. Nichtsdestotrotz ist durch diese relative Unvorhersehbarkeit der Verkehrssituation eine größere Übereinstimmung mit der Realität vorhanden, da es in realen Kontrollsituationen auch zu Verspätungen oder Flugausfällen kommen kann.

Die unstandardisierten Szenarien führten weiterhin zur *Entstehung unterschiedlich komplexer Situationen*. Durch ein verändertes Auftauchen von Flugzeugen können bei einigen Versuchspersonen Konfliktsituationen entstehen, die bei anderen Versuchspersonen nicht entstanden. Diese Situationen können zu höherem kognitiven Aufwand und damit zu veränderten Bewertungen der Beanspruchung führen. Verschieden komplexe Verkehrssituationen könnten aber auch bereits durch ein verändertes Verhalten der Versuchspersonen während der Szenarien entstehen. Dieses Problem besteht aufgrund der Konfundierung der Variablen „Verkehrsaufkommen“ und „Anzahl kritischer Situationen“. Daher wäre als ein besseres Maß für die Belastung die Anzahl der zu treffenden Entscheidungen oder die Anzahl der kritischen Situationen heranzuziehen. Der Zusammenhang zwischen der Schwere der Entscheidung und der mentalen Beanspruchung ist allerdings im Modell noch nicht abgebildet. Es ist unter Umständen sinnvoll, nur sehr kurze Szenarien zu verwenden (etwa eine Minute), die eine bestimmte Anzahl Entscheidungen erfordern, wobei im Anschluss an das Szenario die Beanspruchung bewertet wird.

Mit Hilfe *korrelativer Maße* als auch mit Hilfe des *visuellen Vergleiches der Verläufe* konnte ein Nachweis der Modellgüte erbracht werden. Als problematisch stellte sich hierbei allerdings die *Aussagekraft von Maßen der Abweichung vom exakten Punkt* heraus. Das Konstrukt mentale Beanspruchung, welches im Zentrum der Untersuchungen stand, ist eine hochgradig subjektive

Variable. Zwar sollen mit dem Modellansatz sowohl externe als auch personenspezifische Variablen betrachtet werden. Im vorliegenden basalen Modell wurden individuelle Unterschiede allerdings nur sehr vereinfacht (anhand der maximalen Arbeitsgedächtniskapazität) betrachtet. Der Vergleich absoluter Abweichungen der subjektiven Größe sollte daher mit Vorsicht geschehen, da die festgelegte Maximalkapazität des Arbeitsgedächtnisses ebenso willkürlich sein kann. Ferner besteht das Problem, dass die Erfassung der absoluten Ausprägung mentaler Beanspruchung mit Hilfe konstruierter Skalen recht willkürlich ist. Es ist schwer zu sagen, welches Gefühl der Beanspruchung die Versuchsperson mit einem Wert von 50 auf der RSME-Skala verband. Eine andere Versuchsperson könnte dieselbe Belastung mit einem anderen Beanspruchungswert einschätzen.

Hierbei besteht vor allem auch das Problem, dass die *Skalen nicht äquidistant* sind, das bedeutet, dass kleinere Belastungsunterschiede im unteren Bereich der Skala bereits zu Unterschieden in den Beanspruchungsratings führen können, während im oberen Bereich sehr viel größere Belastungsunterschiede nötig sind, um das Beanspruchungsratig zu verändern. Ferner besteht der Trend, dass mehrheitlich im unteren bis mittleren Bereich der Skala angekreuzt und nur selten der obere Bereich genutzt wird. Es konnte ebenfalls gezeigt werden, dass unterschiedliche Erfassungsmethoden zu unterschiedlichen Zeitpunkten zu unterschiedlich hohen Beanspruchungsbewertungen führen können. So wurden mit dem NASA-TLX im Anschluss an ein Szenario absolut gesehen viel höhere Bewertungen abgegeben, als mit dem RSME während der Szenarien. Es gilt hierbei zu prüfen, ob diese Unterschiede durch den Messzeitpunkt (im Nachhinein vs. während) oder durch die unterschiedliche Skala zustande kamen.

Ein weiteres Problem ergibt sich, will man die *Skala der subjektiven Maße und des Models ineinander umformen* d.h. sie auf gleiche Einheiten bzw. Längen Transformieren (z.B. Prozentwerte). Die Anwendung einer subjektiven Skala, welche ebenso viele Stufen, wie die Modellskala hat, könnte hier Abhilfe schaffen. Für die Annahme einer zehnstufigen Modellskala (zehn chunks maximal im Arbeitsgedächtnis speicherbar) wäre die Anwendung der zehnstufigen Skala des NASA-TLX denkbar. Nichtsdestotrotz konnte auch bei der RSME-Skala durch Umformung der absoluten Werte in Prozent-Werte ein Vergleich mit der Modellskala erreicht werden. Das Problem der nicht gegebenen Äquidistanz bleibt allerdings für alle subjektiven Skalen bestehen. Die Modellskala kann hingegen als äquidistant angesehen werden. Es gilt demnach zu prüfen, inwiefern beispielsweise durch die Anpassung der Modellskala auf die verbalen Anker der RSME-Skala eine bessere Übereinstimmung der Skalen zu erreichen ist.

Die mehrheitliche *Anwendung nonparametrischer Verfahren und Rangkorrelationskoeffizienten* ergab sich ebenso aus der Beschaffenheit der Daten. Da die subjektiven Beanspruchungsdaten nur als ordinalskaliert betrachtet werden können, sind andere als nonparametrische Verfahren

nicht zulässig. Es existieren zahlreiche Beträge darüber, dass beispielsweise die ANOVA robust gegenüber moderaten Verletzungen der Voraussetzungen (wie Normalverteilung, Schiefe und Kurtosis) ist (DeCarlo, 1997; Glass, Peckham, & Sanders, 1972; Harwell, Rubinstein, Hayes, & Olds, 1992), das Skalenniveau gehört allerdings nicht dazu. Dennoch existieren zahlreiche Studien, in denen parametrische Verfahren auf Beanspruchungsdaten angewendet werden, die mit subjektiven Skalen wie NASA-TLX oder RSME erfasst wurden. Dies geht sogar soweit, dass Daten, die über Likert-Skalen erhoben wurden und somit eindeutig nur Ordinalskalenniveau besitzen, mit parametrischen Verfahren ausgewertet werden. Vor allem beim NASA-TLX aber auch beim RSME mag zwar der Eindruck einer Intervallskalierung vermittelt werden, sodass nicht auf den ersten Blick eindeutig ist, welches Skalenniveau tatsächlich vorliegt. Bei zweifelhafter Kardinalskalenqualität sind allerdings nach Bortz und Lienert (2008) immer nonparametrische Verfahren anzuwenden. Wenn nicht davon ausgegangen werden kann, dass *„äquidistante Zahlenabstände auf einer Skala gleiche Merkmalsunterschiede abbilden“* (Bortz & Lienert, 2008, S. 28), wie es sowohl bei der NASA-TLX als auch der RSME-Skala der Fall ist, kann für diese Erfassungsmethoden nur ordinales Skalenniveau angenommen werden. Da die Teststärke nonparametrischer Verfahren gemeinhin etwas geringer ist als die parametrischer Verfahren, kann es passieren, dass kleinere Effekte nicht aufgedeckt werden. Dies gilt vor allem hinsichtlich der Testung der Nullhypothesen und der Abhängigkeit der Fehlerzahl vom d2-R Ergebnis. Ähnliches gilt für die Berechnung von Rangkorrelationskoeffizienten. Gemeinhin sind die Rangkorrelationskoeffizienten Spearman’s ρ aber vor allem Kendall’s τ konservativer als Pearson’s r, wodurch geringere Korrelationskoeffizienten zustande kommen. Es wurde allerdings als wichtiger angesehen, nur dann parametrische Verfahren anzuwenden, wenn die notwendigen Voraussetzungen erfüllt waren (so bei der Herzrate und den Modelldaten geschehen).

Die Schwierigkeiten mit Hilfe der *Herzrate* sensitiv zwischen zwei Belastungsbedingungen unterscheiden zu können bestand auch in der vorliegenden Studie. Es konnte lediglich gezeigt werden, dass die Belastungsbedingungen von der Baseline-Bedingung unterschieden werden konnten. Es ist daher nicht zielführend, die Herzrate als Vergleichsmaß für die Überprüfung der Modellgüte weiterzuverwenden, da auch größere Belastungsunterschiede, nicht aufgedeckt werden konnten. Es gilt daher zu prüfen, ob sich andere physiologische Methoden hierbei als sensitiver und damit geeigneter erweisen.

Der Nachweis der Gültigkeit des Modells konnte durch Zusammenhangsmaße sowohl für die mittlere Beanspruchungshöhe innerhalb der Szenarien, als auch der einzelnen Versuchspersonen erbracht werden. Ferner konnte durch explorative Untersuchungen gezeigt werden, dass auch die absolute mittlere Beanspruchungshöhe innerhalb der Szenarien gut abgebildet werden konnte.

8 AUSBLICK

Die Entwicklung und Validierung des hier vorgestellten Modells konnte in vielerlei Hinsicht zum Erkenntnisgewinn beitragen. Einerseits erfordert es eine intensive Beschäftigung mit der Theorie mentaler Beanspruchung und der Generierung einer adäquaten Möglichkeit zur Modellierung und leistet damit einen Beitrag zum theoretischen Erkenntnisgewinn hinsichtlich des Konstruktes. Andererseits konnte durch die Übertragung der Theorie auf einen Anwendungsfall ein erster Schritt getan werden, mentale Beanspruchung von Towerlotsen zu quantifizieren, zu replizieren und sogar vorherzusagen. Nichtsdestotrotz ist es von großer Bedeutung, zunächst eine einheitliche Definition mentaler Beanspruchung vorzulegen, um einheitliche Standards zu schaffen. Es ist allerdings schwer zu sagen, ob es jemals möglich sein wird, eine einheitliche Definition aufzustellen: *„Is an agreed definition of workload ever possible?“* (Brooker, 2003, S. 4).

Ebenso problematisch wie die einheitliche theoretische Fundierung mentaler Beanspruchung ist ihre Erfassung. Da mentale Beanspruchung keiner direkten Erfassung zugänglich ist, ist es schwierig, über Erfassungsmethoden Rückschlüsse auf die Beanspruchung zu ziehen. Alle Erfassungsmethoden haben Schwachstellen, die den direkten Vergleich zwischen den mit diesen Methoden erfassten Daten und den Modelldaten stark erschweren. So ist die Herzrate zwar als relativ objektiv zu bezeichnen, sie ist allerdings nicht sensitiv genug. Subjektive Fragebogenmethoden und Skalen zur Erfassung haben einerseits das Problem, dass kontinuierliche Messung ohne Interferenz praktisch nicht möglich ist. Andererseits bestehen bereits große Probleme in der Reliabilität und Validität der Messinstrumente selbst, so dass es schwer ist eine geeignete Vergleichsmethodik zu finden. So bleibt die Frage, ob eventuell andere Erfassungsmethoden, wie beispielsweise Pupillometrie, EEG oder sogar Nahinfrarot-spektroskopie wie sie z.B. Ayaz et al. (2010) verwenden, geeignetere Vergleichskriterien sein können oder ob letztendlich sogar die gesamte Problematik der Erfassung mentaler Beanspruchung ad absurdum zu führen ist (Schmidtke, 2002; vgl. auch Dekker & Hollnagel, 2004; Dekker & Woods, 2002; Hollnagel & Woods, 2005).

In beiden Fällen kann nur von ersten Schritten gesprochen werden. Die Theorie mentaler Beanspruchung ist immer noch recht vage und auch wenn durch die Integration der beiden herausragenden Modellvorstellungen ein Schritt in diese Richtung getan wurde, liegt bis heute keine einheitliche Modellvorstellung vor. In dieser Arbeit wurde mentale Beanspruchung durch die Begrenzungen der perzeptiv-kognitiven Verarbeitungsstufe, vor allem des Arbeitsgedächtnisses, gesehen. Dabei wurde die Begrenzung der Perzeption zwar mit modelliert, jedoch bei der Quantifizierung zunächst nicht mit betrachtet. Hierbei ist die Frage essentiell, ob und inwieweit die Begrenzungen des perzeptiven Apparates zum subjektiven

Gefühl des Angestrengtseins und damit zum Empfinden mentaler Beanspruchung beitragen und wie diese durch Interaktion mit dem Arbeitsgedächtnis zustande kommt. Dies gilt insbesondere, wenn man dem Modell die Fähigkeit verleihen möchte, Verkehrssituationen konstant überwachen zu können und nicht nur punktuell Informationen über die Verkehrssituation abrufen zu können. Dies ist vor allem auch dann wichtig, wenn die Aufmerksamkeitsverteilung von Towerlotsen modellhaft beschrieben werden soll (vgl. Manske, von Schlippe, et al., 2013).

Das gilt vor allem auch hinsichtlich einer Erweiterung des Prozesses der Entscheidungsfindung. Im vorliegenden Modell wurde lediglich auf Grundlage der aktuellen Verkehrssituation entschieden, ob eine Freigabe gegeben werden kann oder nicht. Es geht aber bereits aus der Aufgabenanalyse hervor, dass Lotsen die aktuelle Verkehrssituation in die Zukunft projizieren, diese also „mental simulieren" (Klein & Crandall, 1995; Klein et al., 2006; vgl. auch Endsley, 1995) um eventuelle Konflikte vorhersehen zu können und dementsprechend prospektiv ihre Freigaben erteilen zu können. Smieszek und Joeres (2013) greifen diese Problematik auf und identifizieren auch hier, dass weitere Grundlagenforschung nötig ist, einerseits hinsichtlich der kognitiven Mechanismen bei solchen mentalen Simulations- und Zeitschätzprozessen (vgl. Peterken, Brown, & Bowman, 1991; Rußwinkel et al., 2011; Zakay & Block, 1997), andererseits hinsichtlich des damit verbundenen Bedarfs an kognitiven Ressourcen und des Umfanges der dafür benötigten Chunks.

Es ist ferner davon auszugehen, dass die einfache Modellvorstellung eines Single-Limited-Capacity Systems nicht mehr ausreichend ist, um komplexere kognitive Prozesse zu beschreiben. Neuere Theorien des Arbeitsgedächtnisses gehen zwar von einem Multi-Komponenten-Arbeitsgedächtnis aus, bei dem verschiedene Komponenten relativ unabhängig für verschiedene Kodierungen von Informationen arbeiten, dennoch konnte das Zusammenspiel dieser Komponenten bisher nicht geklärt werden. Hierbei ist es vor allem schwierig, die Kapazitätsbegrenzungen der einzelnen Komponenten zu quantifizieren und exakt zuzuordnen. Im Zuge einer Modellerweiterung erscheint es dennoch sinnvoll, aktuelle Theorien des Arbeitsgedächtnisses zu berücksichtigen, wobei auch hier darauf hinzuweisen ist, dass, bevor eine solche Modellerweiterung stattfinden kann, einschlägige Grundlagenforschung notwendig ist, um die aufgeworfenen Fragenstellungen beantworten zu können.

Ebenso wie das Modell verwendet wurde, um mentale Beanspruchung zu quantifizieren, so ist es auch vorstellbar, das Modell einzusetzen, um zuverlässig verschiedene andere Beanspruchungsarten zu erfassen. So kann das motorische Modul beispielsweise verwendet werden, um motorische Beanspruchung zu erheben, ebenso wie das Working-Memory-Modul verwendet wurde, um mentale Beanspruchung zu erfassen.

Für eine Erweiterung der praktischen Anwendungsmöglichkeiten des Modells ist durch dessen modularen Aufbau beispielsweise die Erweiterung auf zwei Lotsen möglich, um so ein realistischeres Abbild der Rollenverteilung im Tower zu erhalten und auch die Interaktionen zwischen den beiden Arbeitspositionen mit in Untersuchungen einfließen lassen zu können. Letztendlich wäre damit auch denkbar, Phänomene wie Teamwork oder Shared Cognition (Hopp & Hayne, 2002; Kiekel & Cooke, 2004) zu untersuchen.

Ebenso ist die Erweiterung des Modells hinsichtlich der sogenannten Remote-Control (Fürstenau et al., 2009; Öhme & Schulz-Rückert, 2010), also der Bearbeitung zweier kleinerer Flughäfen durch einen einzelnen Lotsen, möglich, um Untersuchungen hinsichtlich mentaler Beanspruchung von Lotsen bei der Bearbeitung mehrerer Flughafen zu ermöglichen.

Im Hinblick auf die Verwendung farbiger Petrinetze zur Beschreibung kognitiver Prozesse und menschlichen Verhaltens konnte ein weiterer Schritt getan werden. Es wurde nicht nur formale Abläufe beschrieben, sondern auch kognitive Prozesse, Ressourcen und Strukturen. Aus den Erfahrungen dieser Arbeit ist auch eine kombinierte Verwendung farbiger Petrinetze und kognitiver Architekturen, wie ACT-R, denkbar. Hierbei könnten farbige Petrinetze dazu dienen den Prozess abzubilden, die kognitive Architektur liefert die nötigen Voraussetzungen zur Beschreibung menschlicher Kognition. Prinzipiell kann es sinnvoll sein, nur bestimmte Funktionen, die sehr detailliert abgebildet werden sollen, an die kognitive Architektur auszulagern, die restlichen Vorgänge aber im Petrinetz zu belassen. Darüber hinaus ist auch die Erstellung einer kognitiven Architektur auf der Basis farbiger Petrinetze denkbar, wie sie bereits Werther (2006b) in Ansätzen vorgestellt.

Mit der hier getätigte Modellierungs- und Validierungsarbeit hinsichtlich einer hoch brisanten Fragestellung, sollte ein Anstoß gegeben werden, den Bedarf an der weiteren Forschungsarbeit zu erkennen, sich den weiteren Herausforderung zu stellen und sich dazu ermutigt fühlen sich der offenen Fragestellungen anzunehmen und Lösungen durch die engagierte Zusammenarbeit von vielerlei Seiten in Angriff zu nehmen.

LITERATURVERZEICHNIS

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*(3), 288–318.

Ackerman, P. L. (1990). A correlational analysis of skill specifity: learning abilities, and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 883–901.

Adams, J. A. (1989). *Human factors engineering*. New York: Macmillan.

Alexander, J. R., Alley, V. L., Ammerman, H. L., Fairhurst, W. S., Hostetler, C. M., Jones, G. W., & Rainey, C. L. (1989). *FAA Air Traffic Control Operations Concepts Volume 7: ATCT Tower Controllers*. Oklahoma City: Federal Aviation Administration.

Alexander, J. R., Alley, V. L., Ammerman, H. L., Hostetler, C. M., & Jones, G. W. (1988). FAA Air Traffic Control Operations Concepts Volume 3: ISSS En Route Controllers. Oklahoma City: Federal Aviation Administration.

Ammerman, H. L., Becker, E. S., Bergen, L. G., Claussen, C. A., Davies, D. K., Inman, E. E., & Jones, G. W. (1988). *FAA Air Traffic Control Operations Concepts Volume 5: ATCT/TCCC Tower Controllers*. Oklahoma City: Federal Aviation Administration.

Ammerman, H. L., Fairhurst, W. S., Hostetler, C. M., & Jones, G. W. (1988). FAA Air Traffic Control Operations Concepts Volume 1: ATC Background and Analysis Methodology. Oklahoma City: Federal Aviation Administration.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Anderson, J. R., & Lebiere, C. (1998). *The atomic conponents of thought*. Mahwah, NJ: Lawrence Erlbaum.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: activation limitations on retrieval. *Cognitive Psychology*, *30*(3), 221–56. doi:10.1006/cogp.1996.0007

Annett, J. (2002). Subjective rating scales: science or art ? *Ergonomics*, *45*(14), 966–987.

Athènes, S., Averty, P., Puechmorel, S., Delahaye, D., & Collet, C. (2002). ATC Complexity and Controller Workload: Trying to Bridge the Gap. *Artificial Intelligence*, (56), 56–60.

Ayaz, H., Willems, B., Bunce, S., Shewokis, P. A., Izzetoglu, K., Hah, S., … Onaral, B. (2010). Cognitive Workload Assessment of Air Traffic Controllers Using Optical Brain Imaging Sensors. In T. Marek, W. Karwowski, & V. Rice (Eds.), *Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations* (pp. 21–32). CRC Press.

Baddeley, A. D. (1976). *The Psychology of Memory*. New York: Basic Books.

Baddeley, A. D. (1986). *Working Memory*. Oxford, UK: Oxford University Press.

Baddeley, A. D. (1990). *Human Memory: Theory and Practice*. Hove: Lawrence Erlbaum Associates.

Baddeley, A. D. (1994). The Magical Number Seven: Still Magic After All These Years?, *101*(2), 353–356.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Baddeley, A. D., Grant, W., Wight, E., & Thomson, N. (1975). Imagery and visual working memory. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and Performance V* (pp. 205–217). London: Academic Press.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation, Vol. 8*. New York: Academic Press.

Bainbridge, L. (1975). Working Memory in air traffic control. Unpublished note. Retrieved October 30, 2012, from http://www.bainbrdg.demon.co.uk/Papers/WMemory.html

Baumgarten, B. (1996). *Petri-Netze: Grundlagen und Anwendungen* (2nd ed.). Heidelberg, Berlin, Oxford: Spektrum, Akademischer Verlag.

Beatty, J. (1982). Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin*, *91*(2), 276–292.

Bentley, R., Hughes, J. A., Randall, D., Rodden, T., Sawyer, P., Shapiro, D., & Sommerville, I. (1992). Ethnographically-informed systems design for air traffic control. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work* (pp. 123–129). ACM Press.

Bergé, P. (2005). *Initial EMMA Air-Ground Operational Service and Environmental Description (OSED Initial), Document No. D1.3.1*. EUROCONTROL.

Berndtsson, J., & Normark, M. (1999). The Coordinative Functions of Flight Strips: Air Traffic Control Work Revisited. In *Proceedings of the international ACM SIGGROUP conference on supporting group work* (pp. 101–110). Phoenix, Arizona.

Bisseret, A. (1970). A Mémoire opérationelle et structure du travail. *Bulletin de Psychologie*, *24*, 280–294.

Bisseret, A. (1971). Analysis of mental processes involved in air traffic control. *Ergonomics*, *14*(5), 565–570.

Bortz, J. (2005). *Statistik* (6th ed.). Berlin: Springer.

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler* (4th ed.). Heidelberg: Springer.

Bortz, J., & Lienert, G. A. (2008). *Kurzgefasste Statistik für die Klinische Forschung*. Heidelberg: Springer.

Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7th ed.). Berlin, Heidelberg: Springer.

Bothell, D. (2007). Act-R 6.0 Reference Manual - Working Draft. Retrieved September 23, 2013, from http://act-r.psy.cmu.edu/actr6/reference-manual.pdf

Boucsein, W. (1991). Arbeitspsychologische Beanspruchungsforschung heute - eine Herausforderung für die Psychophysiologie. *Psychologische Rundschau*, *42*, 129–144.

Brickenkamp, R., Schmidt-Atzert, L., & Liepmann, D. (2010). *d2-R: Test d2-Revision Aufmerksamkeits- und Konzentrationstest*. Göttingen: Hogrefe. Retrieved from http://www.testzentrale.de/programm/media/downloads/0141301/0141301.pdf

Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, *47*, 191–196.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.

Broadbent, D. E. (1975). The magical number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long term memory* (pp. 3–18). Bristol: John Wiley & Sons.

Brooker, P. (2003). Control workload, airspace capacity and future systems. *Human Factors and Aerospace Safety*, *3*(1), 1–23.

Brookhuis, K. A., & De Waard, D. (2001). Assessment of driver's workload: Performance and subjective and physiological indexes. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue*. Mahwah, NJ: Lawrence Erlbaum.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, *42*(3), 361–77.

Bub, W., & Lugner, P. (1992). Systematik der Modellbildung. In Verein Deutscher Ingenieure (Ed.), *Modellbildung für Regelung und Simulation* (pp. 1–43). Düsseldorf: VDI-Verlag.

Buck, S., Biemans, M. C. H., Hilburn, B. G., & Van Woerkom, P. T. L. M. (1996). *Synthesis of functions (NLR Report TR-970545L)*. Amsterdam: NLR.

Büttner, P. (2010). "Hello Java!" Linking ACT-R 6 with a Java simulation. In D. D. Salvucci & G. Gunzelmann (Eds.), *Proceedings of the 10th International Conference on Cognitive Modeling* (pp. 289–290). Philadelphia: Drexel University.

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and raw task load index (TLX) correlations: are paired comparisons necessary? In A. Mital (Ed.), *Advances in industrial ergonomics and safety, I* (pp. 481–485). London: Taylor & Francis.

Cacciabue, P. C. (1998a). Modelling and simulation of human behaviour for safety analysis and control of complex systems. *Safety Science*, *28*(2), 97–110.

Cacciabue, P. C. (1998b). *Modelling and Simulation of Human Behaviour in System Control*. London: Springer.

Cacciabue, P. C., & Hollnagel, E. (1995). Simulation of Cognition: Applications. In J.-M. Hoc, P. C. Cacciabue, & E. Hollnagel (Eds.), *Expertise and Technology: Cognition & Human-Computer Cooperation* (pp. 43–54). Hillsdale, New Jersey: Lawrence Erlbaum Associates Ltd.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Card, S. K., Moran, T. P., & Newell, A. (1986). The Model Human Processor - An Engineering Model of Human Performance. In K. R. Boff, L. Kaufmann, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance, Vol. 2: Cognitive Processes and Performance* (pp. 1–35).

Casali, J., & Wierwille, W. (1983). A comparison of rating scale, secondary-task, physiological and primary task workload estimation techniques in simulated flight task emphazizing communications load. *Human Factors*, *25*, 623–642.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81.

Costa, G. (1993). Evaluation of workload in air traffic controllers. *Ergonomics*, *36*(9), 1111–1120.

Courboulay, M., & Kahn, J. (1996). *ATHOS CDG Airport Tower Controller Task Analysis, WP3, ATHOS/SOF-TEC-W3-006-R1, TR 1005*. EUROCONTROL.

Cowan, N. (1998). Visual and auditory working memory capacity. *Trends in Cognitive Sciences*, *2*(3), 77–78.

Cowan, N. (1999). An Embedded-processes Model of Working Memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory* (pp. 62–101). Cambridge: Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87–114.

Cowan, N. (2010). The Magical Mystery Four: How is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, *19*(1), 51–57.

Cox, M. (1994a). *Task Analysis of selected operating Positions within UK Air Traffic Control Volume 1: Main Report*. Farnborough, Hampshire: Royal Air Force Institute of Aviation Medicine.

Cox, M. (1994b). *Task Analysis of selected operating Positions within UK Air Traffic Control Volume 2: Appendices*. Farnborough, Hampshire: Royal Air Force Institute of Aviation Medicine.

CPN-Tools. (n.d.). The CPN-Tools Webpage. Retrieved from http://cpntools.org/

Crandall, B., Klein, G., & Hoffman, R. R. (2006). *Working Minds: A Practioneer's Guide to Cognitive Task Analysis*. Cambridge: MIT Press.

Cybenko, G., & Brewington, B. (1999). The Foundations of Information Push and Pull. In G. Cybenko, D. P. O'Leary, & J. Rissanen (Eds.), *The Mathematics of Information Coding, Extraction and Distribution (The IMA Volumes in Mathematics and its Applications Volume 107)* (pp. 9–30). New York: Springer.

Daams, J., Blom, H. A. P., & Nijhuis, H. B. (2001). *Modelling Human Reliability in Air Traffic Management (NLR-TP-2001-629)*. Amsterdam, The Netherlands: National Aerospace Laboratory NLR.

Daneman, M., & Carpenter, P. A. (1980). Individual Differences in Working Memory and Reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

David, R., & Alla, H. (2010). *Discrete, Continuous, and Hybrid Petri Nets*. Berlin, Heidelberg: Springer.

Davidrajuh, R., & Lin, B. (2011). Exploring airport traffic capability using Petri net based model. *Expert Systems with Applications*, *38*(9), 10923–10931.

De Groot, A. D. (1965). *Thought and choice in chess*. Mouton: The Hauge.

De Waard, D. (1996). *The Measurement of Drivers ' Mental Workload*. Doctoral Dissertation: University of Groningen.

De Waard, D., Jessurun, M., Steyvers, F. J., & Ragatt, P. T. F. (1995). Effect of road layout and road environment on driving performance, drivers' physiology and road appreciation. *Ergonomics*, *38*(7), 1395–1407.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, *2*(3), 292–307.

Dekker, S. W. A., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work*, *6*(2), 79–86.

Dekker, S. W. A., & Woods, D. D. (2002). MABA-MABA or Abracadabra? Progress on Human-Automation Co-ordination. *Cognition, Technology & Work*, *4*(4), 240–244.

Desel, J., & Juhás, G. (2001). What is a Petri Net? - Informal Answers for the Informed Reader. In H. Ehrig & G. Juhás (Eds.), *Lecture Notes in Computer Science 2128* (pp. 1–25). Berlin: Springer.

Desmond, P. A., & Hoyes, T. W. (1996). Workload variation, intrinsic risk and utility in a simulated air traffic control task: Evidence for compensatory effects. *Saftey Science*, *22*(1), 87–101.

DFS. (2002). *Betriebsanweisung Flugverkehrskontrolle*. Offenbach am Main: Deutsche Flugsicherung.

DFS. (2006). *Betriebsanweisung Flugverkehrskontrolle*. Offenbach am Main: Deutsche Flugsicherung.

DFS. (2013). *Air Traffic Statistics - Annual Summary 2013*. Offenbach am Main: Deutsche Flugsicherung.

Dittmann, A., Kallus, K. W., & Van Damme, D. (2000). *Integrated Task and Job Analysis of Air Traffic Controllers - Phase 3: Baseline Reference of Air Traffic Controller Tasks and Cognitive Processes in the ECAC Area*. EUROCONTROL.

Dörner, D. (2000). *Die Logik des Mißlingens - Strategisches Denken in komplexen Situationen* (13th ed.). Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag GmbH.

Dörner, D., & Schaub, H. (1995). Handeln in Unbestimmtheit und Komplexität. *Organisationsentwicklung*, *14*, 34–47.

Duffy, E. (1951). The concept of energy mobilization. *Psychological Review*, *58*, 30–40.

Durso, F. T., Batsakes, P. J., Crutchfield, J. M., Braden, J. B., & Manning, C. a. (2004). The Use of Flight Progress Strips While Working Live Traffic: Frequencies, Importance, and Perceived Benefits. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 32–49.

Durso, F. T., & Manning, C. A. (2008). Air Traffic Control. In M. C. Carswell (Ed.), *Reviews of human factors and Ergonomics, Volume 4* (pp. 195–244). Human Factors and Ergonomics Society.

Dzaack, J. (2008). *Analyse Kognitiver Benutzermodelle für die Evaluation von Mensch-Maschine-Systemen*. Doctoral Dissertation: Technische Universität Berlin.

Eggemeier, F. T., & Wilson, G. F. (1991). Performance-based and subjective assessment of workload in multi-task environments. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 217–278). London: Taylor & Francis.

Eilers, K., Nachreiner, F., & Hänecke, K. (1986). Entwicklung und Überprüfung einer Skala zur Erfassung subjektive erlebter Anstrengung. *Zeitschrift Für Arbeitswissenschaft*, *40*(4), 215–224.

Embrey, D., Humphreys, P., Rosa, E., Kirwan, B., & Rea, K. (1984). *SLIM-MAUD: An Approach to Assessing Human Error Probabilities Using Structured Expert Judgment [NUREG/CR-3518]*. New York: Nuclear Regulatory Commission.

Endsley, M. R. (1995). Towards a theory of Situation Awareness in dynamic systems. *Human Factors*, *37*(1), 32–64.

Endsley, M. R., & Rodgers, M. D. (1994). Situation Awareness Information Requirements for en-route Air Traffic Control. Oklahola City: Federal Aviation Administration.

Engell, S. (1992). Modellgüte und Regelgüte. In Verein Deutscher Ingenieure (Ed.), *Modellbildung für Regelung und Simulation* (pp. 189–207). Düsseldorf: VDI Verlag.

Ericsson, K. A., & Chase, W. G. (1982). Exceptional Memory. *American Scientist*, *70*, 607–615.

ETACHIP. (1996). Model for Task and Job Descriptions of Air Traffic Controllers. HUM.ET1.ST10.DEL-01: EUROCONTROL.

Fahrenberg, J. (1983). Psychophysiologische Methodik. In K. J. Groffmann & L. Michel (Eds.), *Enzyklopädie der Psychologie. Psychologische Diagnostik (Bd. 4)* (pp. 1–192). Göttingen: Hogrefe.

Federal Aviation Administration. (2011). *Air Traffic Control*. Federal Aviation Administration.

Fields, B., Amaldi, P., & Tassi, A. (2003). *Representing collaborative work: The Airport as Common Information Space*. Technical Report IDC-TR-2003-003.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*, 381–391.

Flach, J. M. (2008). Mind the Gap: A skeptical view of Macrocognition. In J. M. Schraagan, L. G. Militello, T. Ormerod, & R. Lipshitz (Eds.), *Naturalistic Decision Making and Macrocognition* (pp. 27–40). Aldershot: Ashgate.

Fleid, A. (2009). *Discovering Statistics using SPSS* (3rd ed.). London: SAGE Publications Ltd.

Flemisch, F., Meier, S., Neuhöfer, J., Baltzer, M., Altendorf, E., & Özyurt, E. (2012). Kognitive und kooperative Systeme in der Fahrzeugführung: Selektiver Rückblick über die letzten Dekaden und Spekulation über die Zukunft. In D. Soeffker & A. Kluge (Eds.), *Kognitive Systeme*. Duisburg: DuEPublico, Duisburg-Essen Publication.

Frese, M., & Zapf, D. (1994). Action as the Core of Work Psychology - A German Approach. In H. C. Triandis, M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (2nd ed., pp. 272–340). Palo Alto, CA: Consulting Psychologists Press.

Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., Papenfuß, A., & Kaltenhäuser, S. (2009). Steps Towards the Virtual Tower: Remote Airport Traffic Control Center (RAiCe). In *ENRI International Workshop on ATM/CNS*. Tokyo, Japan.

Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, *42*, 237–228.

Gopher, D., & Donchin, E. (1986). Workload - An examination of the concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 41–1–41–49). New York: Wiley.

Gore, B. F., Hooey, B. L., Foyle, D. C., & Scott-Nash, S. (2008). Meeting the Challenge of Cognitive Human Performance Model Interpretability Through Transparency : MIDAS v5 . x. In *2nd International Conference on Applied Human Factors and Ergonomics*. Las Vegas.

Gray, W. D., John, B. E., & Atwood, M. E. (1992). The Precis of Project Ernestine or An Overview of a Validation of GOMS. In *Proceedings of the CHI 1992* (pp. 307–312).

Green, D. M., & Sweets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Hacker, W. (1986). *Arbeitspsychologie*. Berlin: VEB Deutscher Verlag der Wissenschaften.

Hacker, W. (2005). *Allgemeine Arbeitspsychologie. Psychische Regulation von Arbeitstätigkeiten* (2nd ed.). Bern: Huber.

Hagemann, T. (2000). *Belastung, Beanspruchung und Vigilanz in der Flugsicherung - unter besonderer Berücksichtigung der Towerlotsentätigkeit*. Frankfurt am Main: Lang.

Halbrügge, M. (2013). ACT-CV: Bridging the Gap between Cognitive Models and the Outer World. In E. Brandenburg, L. Doria, A. Gross, T. Günzler, & H. Smieszek (Eds.), *Tagungsband der 10 Berliner Werkstatt Mensch-Maschine-Systeme* (pp. 204–210). Berlin: Universitätsverlag der Technischen Universität Berlin.

Hanson, E. K. S., Schellekens, J. M. H., Veldman, J. B. P., & Mulder, L. J. M. (1993). Psychomotor and cardiovascular consequences of mental effort and noise. *Human Movement Science*, *12*(6), 607–626.

Hart, S. G., & Staveland, L. E. (1986). *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*. Moffett Field, CA: NASA-Ames Research Center.

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing monte carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases. *Journal of Educational and Behavioral Statistics*, *17*(4), 315–339.

Henderson, P. R., Bakal, D. A., & Dunn, B. E. (1990). Cardiovascular response pattern and speech: a study of air traffic controllers. *Psychosomatic Medicine*, *52*, 17–26.

Heuer, H. (1996). Doppeltätigkeiten. In O. Neumann & F. Sanders (Eds.), *Aufmerksamkeit. Enzyklopädie der Psychologie. BandC/II/2* (pp. 163–218). Göttingen: Hogrefe.

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*, 73–93.

Hockey, R. (1984). Varieties of attentional state: The effect of environment. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention1* (pp. 449–484). Orlando: Academic Press.

Hollnagel, E. (1992). Coping, Coupling and Control - The Modelling of Muddling Through. In *2nd Interdisciplinary Workshop on Mental Models*. Robinson College, Cambridge, UK.

Hollnagel, E. (1993a). *Human Reliability Analysis: Context and Control*. London: Academic Press Ltd.

Hollnagel, E. (1993b). Requirements for dynamic modelling of man-machine interaction. *Nuclear Engineering and Design*, *144*, 375–384.

Hollnagel, E. (1998). *Cognitive Reliabiliy and Error Analysis Method - CREAM* (1st ed.). Oxford: Elsevier Science Ltd.

Hollnagel, E. (2009). *The ETTO Principle: Efficiency-Thoroughness Trede-Off – Why things that go right sometimes go wrong*. Farnham: Ashgate.

Hollnagel, E. (2012). The ETTO principle as ETTOing – or Occam ' s Razor redux. In D. De Waard, K. Brookhuis, F. Dehais, C. Weikert, S. Röttger, D. Manzey, … P. Terrier (Eds.), *Human Factors: a view from an integrative perspective (Proceedings HFES Europe Chapter Conference Toulouse)*.

Hollnagel, E., & Woods, D. D. (1983). Cognitive Systems Engineering: New wine in new bottles. *International Journal of Man-Machine Studies*, *18*, 583–600.

Hollnagel, E., & Woods, D. D. (2005). *Joint Cognitive Systems - Foundations of Cognitive Systems Engineering*. Boca Raton: Taylor & Francis.

Hollnagel, E., Woods, D. D., & Leveson, N. G. (2006). *Resilience Engineering - Concepts & Percepts*. Aldershot: Ashgate.

Hope, R. M., Schoelles, M. J., & Gray, W. D. (2013). Connecting ACT-R to the World with JSON over TCP. In R. West & T. Stewart (Eds.), *Proceedings of the 12th International Conference on Cognitive Modeling* (pp. 354–355). Ottawa: Carleton University.

Hopp, P., & Hayne, S. C. (2002). *Literature Review of Shared Cognition*. College of Business, Colorado State University, Fort Collins, CO.

Hörhold, M. (1994). *Zur Psychophysiologie der Belastungsreaktion. Verlaufsanalysen zum Einfluss psychologischer versus physikalischer Situationsmerkmale und psychologischer versus physiologischer Personenmerkmale*. Frankfurt am Main, Berlin, Bern: Lang.

Hörhold, M., & Walschburger, P. (1998). Depressive Störung als Ausdruck misslingender Handlungskontrolle. Überprüfung einer psychophysiologischen Belastungsdiagnostik. *Zeitschrift Für Klinische Psychologie*, *26*(1), 31–37.

Huber, S. (2012). *Optimierung des Funktionsumfanges von Airport Moving Maps durch Analyse von Runway Incursions*. Doctoral Dissertation: Technische Universität Berlin.

Human-Factors-Consult. (2009). *Virtual Control Tower Research Study: Bericht AP1 – Aufgabenanalyse*. Deutsche Flugsicherung.

ICAO4444. (2007). *Air Traffic Management - procedures for air navigation services (Document 444 ATM/501)* (15th ed.). International Civil Aviation Organisation.

ICAO-9426. (1984). *Air traffic services planning manual* (1st ed.). International Civil Aviation Organisation.

Jackson, A. (1989). *The role of the controller in future ATC systems with enhanced information processing capabilities (EEC Report No 224)*.

Jensen, K. (1997a). *Coloured Petri Nets - Basic Concepts, Analysis Methods and Practical Use Volume 1: Basic Concepts* (2nd ed.). Berlin, Heidelberg: Springer.

Jensen, K. (1997b). *Coloured Petri Nets - Basic Concepts, Analysis Methods and Practical Use Volume 2: Analysis Methods* (2nd ed.). Berlin, Heidelberg: Springer.

Jensen, K. (1997c). *Coloured Petri Nets - Basic Concepts, Analysis Methods and Practical Use Volume 3: Practical Use* (2nd ed.). Berlin, Heidelberg: Springer.

Jensen, K. (1998). An introduction to the practical use of coloured Petri Nets. *Lecture Notes in Computer Science Volume*, *1492*, 237–292.

Jensen, K., & Kristensen, L. M. (2009). *Coloured Petri Nets: Modelling and Validation of Concurrent Systems*. Berlin Heidelberg: Springer.

Jensen, K., Kristensen, L. M., & Wells, L. (2007). Coloured Petri Nets and CPN Tools for modelling and validation of concurrent systems. *International Journal on Software Tools for Technology Transfer*, *9*(3-4), 213–254.

Johannsen, G., Moray, N., Pew, R., Rasmussen, J., Sanders, A., & Wickens, C. (1979). Final Report of the Experimental Psychology Group. In N. Moray (Ed.), *Mental Workload (vol. 8)*. New York: Plenum Press.

John, B. E., & Kieras, D. E. (1994). The GOMS Family of Analysis Techniques : Tools for Design and Evaluation. In *ACM Transactions on Computer-Human Interaction*.

Jorna, P. G. A. M. (1992). Spectral analysis of heart rate and psychological state: A review of its validity ans a workload index. *Biological Psychology*, *34*, 237–257.

Jürgensohn, T. (1997). *Hybride Fahrermodelle*. Sinzheim: Pro Universitate.

Jürgensohn, T. (2002). Bedienermodellierung. In K.-P.- Timpe, T. Jürgensohn, & H. Kolrep (Eds.), *Mensch-Maschine-Systemtechnik* (pp. 107–149). Düsseldorf: Symposion.

Jürgensohn, T., Niessen, C., & Leuchter, S. (2002). Bedienermodellierung: Beispiele. In K.-P. Timpe, T. Jürgensohn, & H. Kolrep (Eds.), *Mensch-Maschine-Systemtechnik* (pp. 149–177). Symposion.

Kahnemann, D. (1973). *Attention and Effort*. Englewood Cliffs, N.J.: Prentice Hall.

Kahnemann, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kallus, K. W., Van Damme, D., & Dittmann, A. (1999). *Integrated Task and Job Analysis of Air Traffic Controllers - Phase 2: Task Analysis of En-route Controllers (HUM.ET1.ST01.1000-REP-04)*. EUROCONTROL.

Kantowitz, B. H., & Sorkir, R. D. (1983). *Human factors: Understanding people-system relationships*. New York: Wiley.

Karrer-Gauß, K. (2012). *Prospektive Bewertung von Systemen zur Müdigkeitserkennung: Ableitung von Gestaltungsempfehlungen zur Vermeidung von Risikokompensation aus empirischen Untersuchungen*. Doctoral Dissertation: Technische Universität Berlin.

Kerr, B. A., & Langolf, G. D. (1977). Speed of Aiming Movements. *Quarterly Journal of Experimental Psychology*, *29*, 475–481.

Kiekel, P. A., & Cooke, N. J. (2004). Human Factors Aspects of Team Cognition. In L. V. Kim-Phuong & R. W. Proctor (Eds.), *Handbook of Human Factors in Web Design* (pp. 107–126). Boca: CRC Press.

Kieras, D. E. (1996). *A Guide to GOMS Model Usability Evaluation using NGOMSL*.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, *12*, 391–438.

Kindsmüller, M. C., Leuchter, S., Schulze-Kissing, D., & Urbas, L. (2004). Modellierung und Simulation menschlichen Verhaltens als Methode der Mensch-. *MMI-Interaktiv*, (7), 4–16.

Kirwan, B., & Ainsworth, L. K. (1992). *A Guide to Task Analysis*. London: Taylor & Francis.

Klapp, S. T., Marshburn, E. A., & Lester, P. T. (1983). Short-term memory does not involve the "working memory" of information processing: The demise of a common assumption. *Journal of Experimental Psychology: General*, *112*, 240–264.

Klein, D. E., Klein, H. A., & Klein, G. (2000). Macrocognition: Linking Cognitive Psychology and Cognitive Ergonomics. In *Proceedings of the 5th International Conference on Human Interactions with Complex Systems* (pp. 173–177). Urbana-Champaign: University of Illinois Urbana-Champaign.

Klein, G. (2008). Naturalistic Decision Making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 456–460.

Klein, G., & Crandall, B. W. (1995). The role of mental simulation in naturalistic decision making. In J. M. Flach, P. Hancock, J. Caird, & K. Vicente (Eds.), *The ecology of human-machine systems* (pp. 324–358). Hillsdale, NJ: Lawrence Erlbaum Associates.

Klein, G., Moon, B., & Hoffmann, R. R. (2006). Making sense of sensemaking 2: A macrocognitive model. In *IEEE Intelligent Systems* (pp. 88–92).

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems*, 81–85.

Knuth, D. E. (1997). *The Art of Computer Programming, Vol. 1 - Fundamental Algorithms* (3rd ed.). Amsterdam: Addison-Wesley.

Koros, A., Della Rocco, P. S., Panjwani, G., Ingurgio, V., & D'Arcy, J.-F. (2003). *Complexity in Air Traffic Control Towers: A Flied Study. Part 1. Complexity Factors (DOT/FAA/CT-TN03/14)*. Atlantic City: Federal Aviataion Administration.

Koros, A., Della Rocco, P. S., Panjwani, G., Ingurgio, V., & D'Arcy, J.-F. (2006). *Complexity in Airport Traffic Control Towers: A Field Study. Part 2. Controller Strategies and Information Requirements (DOT/FAA/TC-06/22)*. Atlantic City: Federal Aviataion Administration.

Kosicki, D. (2011). *Der Einfluss von Testszenarien auf die Bewertung von Usability*. Thesis: Philipps-Universität Marburg.

Kovács, A., Németh, E., & Hangos, K. M. (2005). Modeling and Optimization of Runway Traffic Flow Using Coloured Petri Nets. In *International Conference on Control and Automation (ICCA)* (pp. 881–886). Budapest, Hungary.

Kramar, A. F. (1993). Physiological metrics of Mental Workload: A review of recent progress. In *Psychophysiology of Mental Workload (Schriftenreihe der Bundesanstalt für Arbeitsmedizin, Sonderschrift 2)* (pp. 2–34).

Kramar, A. F., & Spinks, J. (1991). Capacity views of human information processing. In J. R. Jennings & M. Coles (Eds.), *Handbook of cognitive psychology: Central and autonomic nervous system approaches* (pp. 179–249). New York: Wiley.

Lacey, J. I. (1967). Somatic response patterning and stress: Some revisions of activation theory. In M. Appley & R. Trumbull (Eds.), *Psychological stress: Issues in research*. New York: Appleton Century Crofts.

Lacey, J. I., & Lacey, B. C. (1978). Two-way communication between the heart and the brain: Significance of time within the cardiac circle. In E. Meyer & J. Brady (Eds.), *Research in the psychology of human behavior*. Baltimore: John Hopkins University Press.

Laird, J. E. (2012). *The Soar Cognitive Architecture*. Massachusetts Institute of Technology.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). SOAR: An Architecture for General Intelligence. *Artifical Intelligence*, *33*, 1–64.

Langley, P., Laird, J. E., & Rogers, S. (2006). *Cognitive Architectures: Research Issues and Challenges, Computational Learning*. Laboratory, CSLI: Stanford University.

Laurig, W., Becker-Biskaborn, G. U., & Reiche, D. (1971). Software problems in analyzing physiological and work study data. *Ergonomics*, *14*, 625–631.

Lebiere, C., & Anderson, J. R. (2001). Multi-Tasking and Cognitive Workload in an ACT-R Model of a Simplified Air Traffic Control Task. In *Proceedings of the Tenth Conference on Computer Generated Forces and Behavioral Representation* (pp. 1–8).

Lee, K., Feron, E., & Pritchett, A. (2007). Air Traffic Ccomplexity: An Input-Output Approach. In *American Control Conference 2007 ACC07* (pp. 474–479).

Lesire, C., & Tessier, C. (2005). Particle Petri Nets for Aircraft Procedure Monitoring Under Uncertainty. In G. Ciardo & P. Darondeau (Eds.), *ICATPN 2005, LNCS 3536* (pp. 329–348). Berlin Heidelberg: Springer.

Leuchter, S. (2009). *Software Engineering Methoden für die Bedienermodellierung in dynamischen Mensch-Maschine-Systemen.* Doctoral Dissertation: Technische Universität Berlin.

Lim, K. Y., & Long, J. B. (1994). *The MUSE Method for Usability Engineering*. Cambridge: Cambridge University press.

Lindblom, C. E. (1959). The science of "muddling through." *Public Administration Quarterly*, *19*, 79–88.

Lindsley, D. B. (1951). Emotions. In S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley.

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and Predicting Mental Workload in En Route Air Traffic Control - Critical Review and Broader Implications. *Human Factors*, *49*(3), 376–399.

Logie, R. H. (1986). Visuo-spatial processing in working memory. *Quarterly Journal of Experimental Psychology A*, *38*, 229–274.

Mackay, W. E. (2000). Is Paper Safer ? The Role of Paper Flight Strips in Air Traffic Control. *Aviation*, *6*(4), 311–340.

Manning, C., Fox, C., & Pfleiderer, E. (2003). Relationships between Measures of Air Traffic Controller Voice Communications, Taskload and Traffic Complexity. In *Proceedings of the 5th USA/Europe Air Traffic Management R&D Seminar*.

Manning, C., Mills, S., Fox, C., Pfleiderer, E., & Mogilka, H. (2001). The Relationship Between Air Traffic Control Communication Events and Measures of Controller Taskload and Workload. In *Proceedings of the 4th USA/Europe Air Traffic Management R&D Seminar*.

Manske, P. (2013). *Generic Airport Petri Net Model (Unveröffentlichter Report)*. Braunschweig: Deutsches Zentrum für Luft- und Raumfahrt.

Manske, P., Smieszek, H., Hasselberg, A., & Möhlenbrink, C. (2013). Entwicklung eines generischen Flughafen-Modells für die effizientere Makrokognitive Modellierung des Mensch-Maschine-Systems der Flughafenverkehrskontrolle mit farbigen Petri-Netzen. In E. Brandenburg, L. Doria, A. Gross, T. Günzler, & H. Smieszek (Eds.), *Proceedings of the 10th Berlin Workshop on Human-Machine Systems: Foundations and Applications of Human-Machine Interaction* (pp. 497–504). Berlin: Universitätsverlag der Technischen Universität Berlin.

Manske, P., von Schlippe, M., Lange, M., & Möhlenbrink, C. (2013). Expertenbefragung: Erfassung der situationsbezogenen Wichtigkeit von Informationsquellen am Towerlotsenarbeitsplatz. In E. Brandenburg, L. Doria, A. Gross, T. Günzler, & H. Smieszek (Eds.), *Tagungsband der 10 Berliner Werkstatt Mensch-Maschine-Systeme* (pp. 490–496). Berlin: Universitätsverlag der Technischen Universität Berlin.

Manzey, D. (1988). *Determinanten bei Doppeltätigkeiten und ressourcentheoretische Modellvorstellung in der kognitiven Psychologie* (DFVL-FB 88.). Köln: Deutsche Forschungsanstalt für Luft- und Raufahrt.

Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung. In F. Rösler (Ed.), *Ergebnisse und Anwendungen der Psychophysiologie (Enzyklopädie der Psychologie), C, Serie I, Bd. 5)* (pp. 799–864). Göttingen: Hogrefe.

Marti, P. (1998). Structured task analysis in complex domains. *Ergonomics*, *41*(11), 1664–1677.

Marti, P. (2000). The Choice of the Unit of Analysis for Modelling Real Work Settings. *Cognition, Technology & Work*, *2*(2), 62–74.

McElree, B. (2001). Working Memory and Focal Attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *27*(3), 817–835.

McKenzie, I. S. (1992). Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human-Computer Interaction*, *7*, 91–193.

Mensen, H. (2004). *Moderne Flugsicherung: Organisation, Verfahren, Technik*. Berlin: Springer.

Merz, W. (2007). *Flugsicherung – was ist das? (Praktikantenhandbuch)*. Deutsche Flugsicherung.

Metalis, S. A. (1991). Heart Period as a Useful Index of Pilot Workload in Commercial Transport Aircraft. *The International Journal of Aviation Psychology*, *1*(2), 107–116.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive control processes and human multiple-task performance: Part 1. Basic Mechanisms. *Psychological Review*, *104*, 3–65.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, *62*, 81–97.

Miller, G. A., Galanter, E. G., & Pribram, K. H. (1960). *Plans and the structure of behaviour*. New York: Rinehart and Winston.

Miller, S. (2001). *Workload Measures*. University of Iowa.

Mogford, R. H., Guttman, J. A., Morrow, S. L., & Kopardekar, P. (1995). *The complexity construct in air traffic control: A review and synthesis of the literature (DOT/FAA/CT-TN95/22)*. Federal Aviataion Administration.

Möhlenbrink, C. (2011). *Modellierung und Analyse von menschlichen Entscheidungsheuristiken mit farbigen Petrinetzen*. Doctoral Dissertation: Technische Universität Braunschweig.

Möhlenbrink, C., & Friedrich, M. (2009). RemoteCenter: Eine Mikrowelt zur Analyse der mentalen Repräsentation von zwei Flughäfen während einer Lotsentätigkeitsaufgabe. In *8. Berliner Werkstatt Mensch-Maschine-Systeme* (pp. 361–366).

Möhlenbrink, C., Manske, P., & Kirlik, A. (2012). An analysis of a ground traffic control decision support system based on the 3-step principle of heuristic decision making. In D. de Waard, K. Brookhuis, F. Dehais, C. Weikert, S. Röttger, D. Manzey, … P. Trrier (Eds.), *Human Factors: a view from an integrative perspective (Proceedings HFES Europe Chapter Conference Toulouse)*.

Moray, N. (1979). Models and Measures of Mental Workload. In N. Moray (Ed.), *Mental Workload* (Vol. 8.). New York: Plenum Press.

Moray, N. (1986). Modeling Cognitive Activities: Human Limitations in Relation to Computer Aids. In E. Hollnagel, G. Mancini, & D. D. Woods (Eds.), *Intelligent Decision Support in Process Environments (NATO Advanced Study Institute F21)* (pp. 273–291). Berlin: Springer.

Muckler, F. A., & Seven, S. A. (1992). Selecting performance measures: "objective" versus "subjective" measurement. *Human Factors*, *34*, 441–455.

Mulder, G. (1980). *The heart and mental effort*. Groningen: University of Groningen.

Myung, J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44*(1), 190–204.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, *26*, 117–163.

Neisser, U. (1979). *Kognition und Wirklichkeit*. Stuttgart: Klett.

Nerdinger, F. W., Blickler, G., & Schaper, N. (2008). *Arbeits- und Organisationspsychologie*. Heidelberg: Springer.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Newton, R. (2002). Parameters dehind "nonparametric" statistics: Kendall's tau, Somers' D and median differences. *The Stata Journal*, *2*(1), 45–64.

Nickel, P., Eilers, K., Seehase, L., & Nachreiner, F. (2002). Zur Reliabilität, Validität, Sensitivität und Diagnostizität von Herzfrequenz- und Herzfrequenzvariabilität als Indikatoren pyschischer Beanspruchung. *Zeitschrift Für Arbeitswissenschaft*, (56), 22–36.

Niessen, C., & Eyferth, K. (2001). A model of the air traffic controller's picture. *Safety Science*, *37*, 187–202.

Niessen, C., Eyferth, K., & Bierwagen, T. (1999). Modelling cognitive processes of experienced air traffic controllers. *Ergonomics*, *42*(11), 1507–1520.

Niessen, C., Leuchter, S., & Eyferth, K. (1998). A Psychological Model of Air Traffic Control and Its Implementation. In *Proceedings of the Second European Conference on Cognitive Modelling (ECCM)* (pp. 104–111). Berlin: Nottingham University Press.

Nolan, M. S. (2011). *Fundamentals of Air Traffic Control* (5th Ed.). Belmont, USA: Brooks/Cole - Thomson Learning.

Norman, D., & Bobrow, D. (1975). On data-limited and ressource-limited processes. *Journal of Cognitive Psychology*, *7*, 44 – 64.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance. Volume II, cognitive processes and performance.* (pp. 42/1–42/49). New York: Wiley.

Oberauer, K., & Bialkova, S. (2009). Accessing information in working memory: can the focus of attention grasp two elements at the same time? *Journal of Experimental Psychology: General*, *138*(1), 64–87.

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity - facets of a cognitive ability construct. *Personality and Individual Differences*, *29*, 1017–1045.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory : Storage, processing, supervision, and coordination. *Intelligence*, *31*, 167–193.

Oberheid, H. (2006). A coloured Petri net model of cooperative arrival planning in air traffic control. In *CPN'06 - Seventh Workshop and Tutorial on Practical Use of Coloured Petri Nets and CPN Tools*. Aarhus, Denmark.

Oberheid, H., & Söffker, D. (2008). Cooperative Arrival Management in Air Traffic Control - A Coloured Petri Net Model of Sequence Planning. In K. M. van Hee & R. Valk (Eds.), *Petri Nets* (pp. 348–367). Berlin, Heidelberg: Springer.

Öhme, A., & Schulz-Rückert, D. (2010). Distant Air Traffic Control for Regional Airports. In F. Vanderhaegen (Ed.), *11th IFAC, IFIP, IFORS, IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems* (pp. 141–145). Université de Valenciennes et du Hainaut Cambrésis, France.

Paul, S., Zografos, K., & Hesselink, H. (2000). *MANTEA Final report MANTEA/ISR-DOC-D83-137-R1, TR 1036*.

Pawlak, W. S., Brinton, C. R., Crouch, K., & Lancaster, K. M. (1996). A Framework for the Evaluation of Air Traffic Control Complexity. In *Proceedings of the AIAA Guidance, Navigation and Control Conference*. San Diego.

Peterken, C., Brown, B., & Bowman, K. (1991). Predicting the future position of a moving target. *Perception*, *20*, 5–16.

Petri, C. A. (1962). *Kommunikation mit Automaten*. Bonn: Institut für Instrumentelle Mathematik, Schriften des IIM Nr. 2.

Pinska, E. (2006). An investigation of the head-up time at tower and ground control positions. In *5th Eurocontrol Innovative Research Workshop* (pp. 81–86).

Pinska, E. (2008). *Analysis of Visual Attention in Augmented Reality Environments for Airport Tower Control*. Doctoral Dissertation: École Pratique des Hautes Études.

Pinska, E., & Bourgois, M. (2005). *Behavioural Analysis of the Tower Controller Activity*. Brétigny-sur-Orge, France.

Pinska, E., & Tijus, C. (2007). Augmented Reality Technology For Control Tower Analysis Of Applicability Based On The Field Study. In *1st CEAS European Air and Space Conference* (pp. 573–580).

Polson, M. C., & Friedman, A. (1988). Task-sharing within and between hemispheres: A multiple-resources approach. *Human Factors*, *30*, 633–643.

Pompanon, C., & Raufaste, E. (2009). The Intervention Trigger Model: Computational Modelling of Air Traffic Control. In *Proceedings of the 2009 Annual Meeting of the Cognitive Science Society* (pp. 2262–2267).

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*, 160–174.

Preim, B., & Dachselt, R. (2010). *Interaktive Systeme - Band 1*. Heidelberg: Springer.

Pribram, M. C., & McGuniess, D. (1975). Arousal, activation and effort in the control of attention. *Psychological Review*, *82*, 116–149.

Rantanen, E. M., & Nunes, A. (2005). Hierarchical Conflict Detection in Air Traffic Control. *The International Journal of Aviation Psychology*, *15*(4), 339–362.

Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernetics*, *13*(3), 257–266.

Ratzer, A. V., Wells, L., Lassen, H. M., Laursen, M., Qvortrup, J. F., Stissing, M. S., … Jensen, K. (2003). CPN Tools for Editing, Simulating, and Analysing Coloured Petri Nets. *Lecture Notes in Computer Science*, *2679*, 450–462.

Raufaste, E. (2006). ATC in ACT-R: A model of conflict detection between planes. In *Proceedings of the thirteenth annual ACT-R workshop* (pp. 102–107). Pittsburgh.

Reason, J. (1991). *Human Error*. Cambridge: Cambridge University Press.

Redding, R. E., & Seamster, T. L. (1994). Cognitive task analysis in air traffic controller and aviation crew training. In N. Johnston, N. McDonald, & R. Fuller (Eds.), *Aviation Psychology in Practice* (pp. 190–222). Aldershot: Averbury Technical.

Reisig, W. (1982). *Petri Nets: An Introduction*. Berlin, Heidelberg, New York: Springer.

Reisig, W. (2010). *Petrinetze: Modellierungstechnik, Analysemethoden, Fallstudien*. Wiesbaden: Vieweg + Teubner.

Ribback, S. (2003). *Psychophysiologische Untersuchung mentaler Beanspruchung in simulierten Mensch-Maschine-Interaktionen*. Doctoral Dissertation: Universität Potsdam.

Richter, P., Wagner, T., Heger, R., & Wiese, G. (1998). Psychophysiological analysis of mental load during driving on rural roads: A quasi-experimentel field study. *Ergonomics (Special Issue: Psychophysiology in Erfonomics)*, *41*(5), 593–609.

Robbins, T. W. (1997). Arousal systems and attentional. *Biological Psychology*, *45*, 57–71.

Robbins, T. W., & Everitt, B. J. (1994). Arousal systems and attention. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 703–720). London: MIT Press.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–367.

Röbig, A., König, C., & Hofmann, T. (2010). Entwicklung eines Low-Cost-Towersimulators zur Evaluation arbeitswissenschaftlicher Fragestellungen. In *USEWARE 2010 - Grundlagen, Methoden, Technologien* (pp. 67–76). Baden-Baden: VDI-Berichte 2009.

Rohmert, W. (1984). Das Belastungs-Beanspruchungs-Konzept. *Zeitschrift Für Arbeitswissenschaft*, *38*, 193–200.

Rohmert, W., & Rutenfranz, J. (1972). *Benutzerhandbuch zu den Untersuchungen zur psychophysiologischen Belastung und Beanspruchung von Fluglotsen*. Institut für Arbeitswissenschaft: Technische Hochschule Darmstadt.

Roscoe, A. H. (1992). Assessing Pilot Workload - Why measure Heart-Rate, HRV and Respiration. *Biological Psychology*, *34*(2-3), 223–228.

Roscoe, A. H. (1993). Heart-Rate as a Psychophysiological Measure for in-Flight Workload Assessment. *Ergonomics*, *36*(9), 1055–1062.

Rose, R. M., & Fogg, L. F. (1993). Definition of a responder. Analysis of behavioral, cardiovascular and endocrine response to varied workload in air traffic controllers. *Psychosomatic Medicine*, *55*, 325–338.

Roske-Hofstrand, R. J., & Murphy, E. D. (1998). Human Information Processing in Air Traffic Control. In M. W. Smolensky & E. S. Stein (Eds.), *Human Factors in Air Traffic Control* (pp. 65–114). San Diego: Academic Press.

Rossi, M., Gonnord, C., Paul, S., Darche, N., Mariano, P., Paggio, R., … Naves, P. (1996). *MANTEA Domain Models, WP3, MANTEA/ALE-TEC-D3.2-014-R4, TR 1036, Telematics Application Programme, (Transport/Air).*

Rouse, W. B., Edwards, S. L., & Hammer, J. M. (1993). Modelling the dynamics of mental workload and human performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, *23*, 1662–1671.

Ruckdeschel, W. (1997). *Modellierung regelbasierten Pilotenverhaltens mit Petrinetzen*. Düsseldorf: VDI Verlag.

Ruckdeschel, W., & Onken, R. (1994). Modelling of Pilot Behaviour Using Petri Nets. In R. Valette (Ed.), *Lecture Notes in Computer Science Vol. 815: Application and Theory of Petri Nets* (pp. 436–453). Berlin Heidelberg: Springer.

Russwinkel, N., Urbas, L., & Thüring, M. (2011). Predicting temporal errors in complex task environments: A computational and experimental approach. *Cognitive Systems Research*, *12*(3-4), 336–354.

Salvucci, D. D., & Lee, F. J. (2003). Simple cognitive modeling in a complex cognitive architecture. In *Proceedings of the conference on Human factors in computing systems - CHI '03*. New York, New York, USA: ACM Press.

Sanders, A. F. (1983). Towards a model of stress and human performance. *Acta Psychologica*, *53*, 61–97.

Sanders, M. S., & McCormick, E. J. (1993). *Human factors in engineering and design*. New York: McGraw-Hill.

Sarris, V., & Reiß, S. (2005). *Kurzer Leitfaden der Experimentalpsychologie*. München: Pearson Studium.

Schmidt, D. K. (1976). On modeling ATC workload and sector capacity. *Journal of Aircraft*, *13*(7), 531–537.

Schmidt, M., Rudolph, M., Werther, B., & Fürstenau, N. (2006). Remote Airport Tower Operation with Augmented Vision Video Panorama HMI. In *2nd International Conference Research in Air Transportation ICRAT* (pp. 221–230). Belgrade, Serbia.

Schmidtke, H. (1976). *Handbuch der Ergonomie*. München: Carl Hanser Verlag.

Schmidtke, H. (2002). Vom Sinn und Unsinn der Messung psychischer Belastung und Beanspruchung. *Zeitschrift Für Arbeitswissenschaft*, (56), 4–9.

Schunn, C. D., & Wallach, D. (2005). Evaluating Goodness-of-Fit in Comparison of Models to Data. In *Psychologie der Kognition: Reden and Vorträge anlässlich der Emeritierung von Werner Tack* (pp. 115–154). Saarbrücken: University of Saarland Press.

Schwer, H., Furedy, J. J., & Heslegrave, R. (1984). Phasic T-wave and heart rate changes as indices of mental effort and task incentive. *Psychophysiology*, *21*(3), 326–333.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psycological Review*, *84*, 127–190.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Ergnonimics*, *69*, 99–118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.

Simon, H. A. (1974). How Big Is Chunk ? *Science*, *183*(4124), 482–488.

Smieszek, H. (2011). Erstellung eines integrativen Prozessmodells der Aufgaben von Rollverkehrslotsen auf Grundlage bestehender Aufgabenanalysen. In *1. interdisziplinären Workshop "Kognitive Systeme": Mensch, Teams, Systeme, Automaten*. Duisburg: Universität Duisburg-Essen.

Smieszek, H., Huber, S., & Jürgensohn, T. (2011). Ein integratives Prozessmodell der Hauptaufgaben von Rollverkehrslotsen als Grundlage zur Modellierung und Simulation. In *9. Berliner Werkstatt Mensch-Maschine-Systeme: Reflexionen und Visionen der Mensch-Maschine-Interaktion* (pp. 429–432). Berlin, Düsseldorf: VDI Verlag.

Smieszek, H., & Joeres, F. (2013). Prospective decision making in a macro-cognitive model of airport traffic control system (MATriCS) based on coloured petri nets. In E. Brandenburg, L. Doria, A. Gross, T. Günzler, & H. Smieszek (Eds.), *Proceedings of the 10th Berlin Workshop on Human-Machine Systems: Foundations and Applications of Human-Machine Interaction* (pp. 505–512). Berlin: Universitätsverlag der Technischen Universität Berlin.

Smieszek, H., & Karl, C. (2012). An approach to cognitive simulation of air traffic controllers based on coloured petri nets. In D. Soeffker & A. Kluge (Eds.), *Kognitive Systeme*. Duisburg: DuEPublico, Duisburg-Essen Publication.

Smieszek, H., Manske, P., Hasselberg, A., Russwinkel, N., & Möhlenbrink, C. (2013). Cognitive Simulation of Limited Working Memory Capacity Applied to an Air Traffic Control Task. In R. West & T. Stewart (Eds.), *Proceedings of the 12th International Conference on Cognitive Modeling* (pp. 227–232). Ottawa, Canada: Carleton University.

Smieszek, H., & Russwinkel, N. (2013). Micro-cognition and macro-cognition: trying to bridge the gap. In E. Brandenburg, L. Doria, A. Gross, T. Günzler, & H. Smieszek (Eds.), *Proceedings of the 10th Berlin Workshop on Human-Machine Systems: Foundations and Applications of Human-Machine Interaction* (pp. 335–341). Berlin: Universitätsverlag der Technischen Universität Berlin.

Smith, R. C. (1980). *Stress, Anxiety and the Air Traffic Control Specialist. Some Conclusions from a Decade of Research*. Oklahoma City: Federal Aviation Administration.

Soede, M., Coeterier, J. F., & Stassen, H. G. (1971). Time analyses of the tasks of approach controllers in ATC. *Ergonomics*, *14*(5), 591–601.

Sperandio, J.-C. (1969). Les variations du partage des taches entre un operater et son assistant, en fonction de la charge de travail du systeme. *Bulletin Du CERP*, *18*, 81–98.

Sperandio, J.-C. (1978). The Regulation of Working Methods as a Function of Workload among Air Traffic Controllers. *Ergonomics*, *21*(3), 195–202.

Staal, M. A. (2004). *Stress, Cognition, and Human Performance: A Literature Review and Conceptual Framework*. Moffett Field, CA: National Aeronautics and Space Administration.

Stein, E. S. (1998). Human Operator Workload in Air Traffic Control. In M. W. Smolensky & E. S. Stein (Eds.), *Human Factors in Air Traffic Control* (pp. 155–183). San Diego: Academic Press.

Stein, E. S., Garland, D. J., & Muller, J. K. (2010). Air-Traffic Controller Memory. In J. A. Wiese, D. V. Hopkin, & D. J. Garland (Eds.), *Handbook of Aviation Human Factors*. Boca Raton: CRC Press.

Stroeve, S. H., Everdij, M. H. C., & Blom, H. A. P. (2011). Studying hazards for resilience modelling in ATM - Mathematical Approach towards Resilience Engineering in ATM (MAREA). In *1st SESAR Innovation Days* (pp. 1–8).

Stubbe, T. (2011). *Kommunikationssysteme in der Flugverkehrskontrolle*. Technische Universität Berlin: Bachelorarbeit.

Stütz, P. (2000). *Adaptive Modellierung des regelbasierten Pilotenverhaltens in Cockpitassistenzsystemen*. Berlin: Dr. Köster.

Sun, R. (2006). The CLARION Cognitive Architecture: Extending Cognitive Modeling to Social Simulation. In *Cognition and Multy-Agent Interaction*. New York: Cambridge University Press.

Svenmarck, P., & Dekker, S. W. A. (2003). Decision support in fighter aircraft: from expert systems to cognitive modelling. *Behaviour & Information Technology*, *22*(3), 175–184.

Swain, A., & Guttman, H. (1983). *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications [NUREG/CR-1278]*. New York: Nuclear Regulatory Commission.

Szpyrka, M. (2006). Analysis of RTCP-nets with reachability graphs. *Fundamenta Informaticae*, *74*(2-3), 375 – 390.

Szpyrka, M. (2008). Modelling and Analysis of Real-time Systems with RTCP-nets. In V. Kordic (Ed.), *Petri Nets: Theory and Applications* (pp. 17–40). Vienna, Austria: I-Tech.

Taatgen, N. A. (2001). A model of individual differences in learning air traffic control. In *Proceedings of the 4th International Conference on Cognitive Modeling* (pp. 361–371).

Taatgen, N. A. (2002). A Model of Individual Differences in Skill Acquisitin in the Kanfer-Ackerman Air Traffic Control Task. *Cognitive Systems Research*, *3*(1), 103–112.

Tattersall, A. J., & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, *39*(5), 740–748.

Tavanti, M. (2006a). *Control Tower Operations: A Literature Review of Task Analysis Studies*. Roma, Italy: EUROCONTROL.

Tavanti, M. (2006b). *Control Tower Operations: Roles Description*. Brétigny-sur-Orge Cedex, France: EUROCONTROL Experimental Centre.

Thonrnhill, W. S. (1995). *The controller capacity computer programm: Measuring controller workload*. Unpublished Report.

Thorngate, W. (1987). Ordinal Pattern Analysis: A Method for Assessing Theory-Data Fit. In W. J. Baker, M. E. Hyland, H. van Rappard, & A. W. Staats (Eds.), *Current Issues in theoretical Psychology* (pp. 345–364). Amsterdam: Elsevier.

Thorngate, W., & Edmonds, B. (2012). Ordinal Pattern Analysis : A simple method for assessing simulation-observation fit (CPM Report No.: CPM-12-216).

Thorngate, W., & Edmonds, B. (2013). Measuring Simulation-Observation Fit: An Introduction to Ordinal Pattern Analysis. *Journal of Artificial Societies and Social Simulation*, *16*(2).

Timpe, K.-P., & Kolrep, H. (2002). Das Mensch-Maschine-System als Interdisziplinärer Gegenstand. In K.-P. Timpe, T. Jürgensohn, & H. Kolrep (Eds.), *Mensch-Maschine-Systemtechnik* (pp. 9–40). Düsseldorf: Symposion.

Todd, P., & Gigerenzer, G. (2001). Putting Naturalistic Decision Making into the Adaptive Toolbox. *Journal of Behavioral Decision Making*, *14*, 381–383.

Treisman, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*, 242–248.

Truitt, T. R. (2005). Implementing electronic flight data in airport traffic control towers. *Proceedings 13th International Symposium Aviation*.

Truitt, T. R. (2006). *Electronic Flight Data in Airport Traffic Control Towers: Literature Review*. Atlantic City: Federal Aviation Administration.

Urbas, L., Heinath, M., Schaft, T., & Leuchter, S. (2007). Die Bewertungsmethode Multitasking GOMS und ihre Anwendung für die Bewertung von In-Vehicle Information Systems. In *Mensch & Computer 2007 Workshopband* (pp. 63–66). Weimar: Verlag der Bauhausuniversität Weimar.

Urbas, L., & Leuchter, S. (2008). Usability Engineering of "In Vehicle Information Systems" With Multi-Tasking GOMS. In *Dachkongress Patientensicherheit und Medizintechnik*.

Van der Linden, M. (1998). The relationships between working memory and long-term memory. *Académie Des Science*, *321*, 175–177.

Vicente, K. J. (1999). *Cognitvie Work Analysis*. Mahwah, NJ: Lawrence Erlbaum.

Vidosavljević, A., & Toŝić, V. (2010). Modeling of turnaround process using petri nets. In *Proceedings (CD) of the 14th ATRS World Conference* (pp. 1–13). Porto, Portugal.

Vingelis, P. J., Schaeffer, E., Stringer, P., Gromelski, S., & Ahmed, B. (1990). *Air Traffic Controller Memory Enhancement: Literature Review and Proposed Memory Aids (DOT/FAA/CT-TN90/38)*. Atlantic City: Federal Aviation Administration Technical Center.

Vogt, J., Hagemann, T., & Kastner, M. (2006). The lmpact of Workload on Heart Rate and Blood Pressure in En-Route and Tower Air Traffic Control. *Journal of Psychophysiology*, *20*(4), 297–314.

Vogt, J., & Kastner, M. (2001). Psychophysiological Monitoring of air traffic controllers: Exploration, simulation, validation. In *Progress in Ambulatory Assessment. Computer-assisted psychological and psychophysiological Methods in Monitoring and Field Studies*. Bern, Göttingen: Hogrefe & Huber Publishers.

Wandmacher, J. (2002). *GOMS-Analysen mit GOMSED*. Darmstadt.

Wastell, D., Sauer, J., & Hockey, R. (2003). Using Micro-worlds in research on distributed Cognition in complex dynamic worlds: A ten Year Retrospective on the Cabin Air Management System (CAMS). In G. C. van der Veer & J. F. Hoorn (Eds.), *Proceedings of the*

9th European Conference on Cognitive Science Approaches to Process Control (pp. 145–150). Amsterdam.

Welford, A. T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, *3*, 189–230.

Welford, A. T. (1968). *Fundamentals of skill*. London: Methuen.

Werther, B. (2006a). Colored Petri net based modeling of airport control processes. In *International Conference on Computational Intelligence for Modeling Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce*.

Werther, B. (2006b). *Kognitive Modellierung mit Farbigen Petrinetzen zur Analyse menschlichen Verhaltens*. Doctoral Dissertation. Technische Universität Braunschweig.

Werther, B., Möhlenbrink, C., & Rudolph, M. (2007). Colored Petri Net Based Formal Airport Control Model for Simulation and Analysis of Airport Control Processes. In V. G. Duffy (Ed.), *Digital Human Modeling* (pp. 1027–1036). Berlin Heidelberg: Springer.

West, R. L., Hancock, E., Somers, S., MacDougall, K., & Jeanson, F. (2013). The Macro Architecture Hypothesis: Applications to Modeling Teamwork, Conflict Resolution, and Literary Analysis. In R. West & T. Stewart (Eds.), *Proceedings of the 12th International Conference on Cognitive Modeling*. Ottawa: Carleton University.

West, R. L., & Pronovost, S. (2009). Modeling SGOMS in ACT-R: Linking Macro- and Microcognition. *Journal of Cognitive Engineering and Decision Making*, *3*(2), 194–207.

Wickens, C. D. (1980). The structure of attentional Ressources. In R. Nickerson (Ed.), *Attention and performance* (pp. 239–257). Hillsdale, NJ: Lawrence Erlbaum.

Wickens, C. D. (1984). Processing Ressources in Attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention*. Orlando: Academic Press.

Wickens, C. D. (1991). Processing Ressources and attention. In D. Damos (Ed.), *Multiple task performance* (pp. 3–34). London: Taylor & Francis.

Wickens, C. D. (1992). *Engineering psychology and human performance*. New York: HarperCollins.

Wickens, C. D. (1999). Aerospace Psychology. In P. A. Hancock (Ed.), *Human Performance and Ergonomics* (2nd ed., pp. 195–242). San Diego: Academic Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*, *3*(2), 157–177.

Wickens, C. D., Helleberg, J., Goh, J., Xu, X., & Horrey, W. J. (2001). *Pilot Task Management: Testing an Attentional Expected Value Model of Visual Scanning*. Aviation Research Lab, Institute of Aviation.

Wickens, C. D., Mavor, A. S., & McGee, J. P. (1997). *Flight to the Future - Human Factors in Air Traffic Control*. Washington D.C.: National Academy Press.

Wickens, C. D., & McCarley, J. S. (2008). *Applied Attention Theory*. Boca Raton: Taylor & Francis.

Wickens, C. D., McCarley, J. S., Alexander, A., Thomas, L., Ambinder, M., & Zheng, S. (2007). Attention-situation awareness (A-SA) model of pilot error. In D. Foyle & B. Hooey (Eds.), *Pilot performance models* (pp. 213–240). Mahwah, NJ: Lawrence Erlbaum.

Wierwille, W., Rahimi, M., & Casali, J. (1985). Evaluation of workload using a simulated flight task in a moving base aircraft simulator. *Human Factors*, *27*, 489–502.

Wilco Publishing. (2008). Tower Simulator. Retrieved July 12, 2013, from http://www.towersimulator.com/

Wittbrodt, N., & Thüring, M. (2010). Identification of Performance Parameters in Complex and Dynamic Work Environments: The Example of Aerodrome Air Traffic Control. In D. de Waard, M. B. Axelsson, B. Peters, & C. Weikert (Eds.), *Human Factors: A system view of human, technology and organisation* (pp. 313–326). Maastricht, Netherlands: Shaker Publishing.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459–482.

Zakay, D., & Block, R. A. (1997). Temporal Cognition. *Current Directions in Psychological Science*, *6*(1), 12–16.

Zeigarnik, B. W. (1927). Das Behalten erledigter und unerledigter Handlungen. In K. Lewin (Ed.), *Untersuchungen zur Handlungs- und Affektpsychologie* (pp. 1–85). Berlin: Psychologisches Institut der Universität Berlin.

Zijlstra, F. R. H., & Van Doorn, L. (1985). *The construction of a scale to measure perceived effort.* Delft, The Netherlands: Department of Philosophy and Social Sciences, Delft University of Technology.

Zsambok, C. R., & Klein, G. (1997). *Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum Associates.